BENEMÉRITA UNIVERSIDAD
AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS
LICENCIATURA EN ACTUARÍA

Análisis de la calendarización y asignación presupuestal de los proyectos de una compañía especializada en la propulsión de motores utilizando simulación Monte Carlo

TESIS PRESENTADA PARA OBTENER EL TÍTULO DE LICENCIADO EN ACTUARÍA

PRESENTA
JOSAFAT SALAS ESCALANTE

DIRECTOR DE TESIS
DR. JOSÉ RAÚL CASTRO ESPARZA

PUEBLA, PUE. OCTUBRE 2017

A mi persona, las razones de mi empeño y cada una de mis versiones, especialmente la que ha encarnado en las proximidades de un año.

A la sabiduría y ejemplo ecuánime de mi padre.

Al prosélito apoyo y compañía de mi madre y hermanos.

Agradecimientos

Ofrezco mi gratitud al Dr. José Raúl Castro Esparza por haberme provisto de un sinnúmero de utensilios y consejos imprescindibles en el desarrollo de este trabajo y mi futura vida profesional.

Doy un reconocimiento a cada uno de los profesores que emplearon su tiempo y conocimiento en la cimentación de mis habilidades.

Finalmente, quiero corresponder a mis amigos y a mi familia por darme virtual auxilio durante mi carrera en medida de sus posibilidades.

Contenido

Capítulo 1

Introducción

La administración de proyectos es un conjunto de técnicas utilizadas para lograr los objetivos y cubrir las necesidades planteados al iniciar un proyecto, y son desarrolladas por los gerentes de proyectos y tomadores de decisiones con el fin de satisfacer las demandas de las instituciones u organizaciones de las que forman parte.

Cada proyecto emprendido involucra incertidumbre y requiere de estrategias de administración que ayuden a enfrentar los riesgos y aprovechar las oportunidades que la incertidumbre conlleva.

La comprensión de los riesgos involucrados en un proyecto brinda a los encargados de los mismos la capacidad de satisfacer en mayor medida las expectativas. La evaluación de los riesgos y la incertidumbre proporciona más información para la toma de decisiones en el desarrollo de los proyectos y el logro de sus metas. La administración de proyectos con información de riesgos agrega valor en varios niveles a los proyectos emprendidos.

La administración de proyectos contemplando riesgos trae consigo ventajas tales como el reconocimiento de la incertidumbre y la posibilidad de obtener pronósticos de posibles resultados, la producción de mejores resultados en los negocios a través de una toma de decisiones más informada, el mejoramiento de la creatividad y la innovación, un mejor control y seguimiento de los proyectos y contribuciones al éxito de los proyectos.

La administración de riesgos es además un componente esencial de la estimación de costos en los proyectos y la programación y administración de los mismos. Como parte integral de la administración de proyectos, la administración de riesgos es una herramienta de uso bastante frecuente.

En el Reino Unido existe un centro de propulsión cuyas tareas consisten en invertir en tecnologías y sus productos asociados, fomentar la colaboración entre las PYME, los fabricantes de vehículos y sus proveedores, preparar a las tecnologías de propulsión avanzadas para ser aplicadas en vehículos dentro y fuera de las carreteras, en otras palabras, convertir las tecnologías en productos, y trabajar con los proveedores para entregar una completa infraestructura de desarrollo de propulsión. Esta institución se encuentra desarrollando proyectos cuyos montos presupuestales y fechas de emprendimiento y entrega de proyectos se encuentran sujetos a factores que involucran incertidumbre. Mediante la administración de proyectos y la administración de riesgos se pretende cumplir los objetivos planteados a continuación.

1.1 Objetivo General

Estructurar un modelo de simulación en la hoja de cálculo que ayude a generar estadísticas de interés para la optimización del presupuesto de los proyectos del Centro Avanzado de Propulsión (CAP).

1.2 Objetivos Específicos

- Desarrollar el modelo de simulación en la hoja de cálculo.

- Estimar el presupuesto gastado en cada subproyecto a través de la variable tiempo por medio de la regresión polinomial.

- Analizar los datos disponibles y definir las distribuciones de probabilidad asociadas a cada variable del modelo.

- Estimar la media y varianza del presupuesto total requerido para los proyectos.

- Hallar el presupuesto necesario para que el riesgo de que sea insuficiente no rebase el porcentaje establecido.

1.3 Alcances

- La base de datos utilizada para el modelo cuenta con información bastante completa, incluyendo los parámetros de las variables modeladas, y las respectivas distribuciones asociadas, por lo que no es necesario realizar ajustes de distribución.

- Los comportamientos de las variables asociadas a las desviaciones de tiempo y presupuesto de cada proyecto son muy similares entre sí.

- La fecha de inicio de gran parte de los proyectos es anterior a la fecha de referencia del análisis, así que se cuenta con una buena cantidad de datos presupuestales históricos, en comparación con la de los valores presupuestales inciertos.

1.4 Limitaciones

- Debido a que se perdió un seguimiento del avance del proyecto, los valores históricos de los datos presupuestales están limitados a una fecha ubicada en el año 2016.

- Dado el inmenso tamaño de la base de datos, el número de variables que se deben simular, y el uso de recursos computacionales que implica el hecho de combinar el software de análisis de riesgo con funciones tales como las macros y recursos de programación, restringe al uso de herramientas propias de la hoja de cálculo.

- El hecho de que existen variaciones en el inicio y fin de los proyectos complica la generalización de las formulas en la hoja de cálculo con funciones de uso común.

Capítulo 2

Variables Aleatorias

Frecuentemente, cuando un experimento es realizado, es de interés principal alguna función del resultado en oposición al resultado mismo. Por ejemplo, al lanzar un par de dados, a menudo, alguien puede interesarse en la suma de los resultados de los dos dados, y realmente, no son de interés los resultados separados de cada dado. Así que, se puede estar interesado en saber que la suma es 7 y puede no ser de importancia que los resultados fueron (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), o (6,1). Además, al lanzar una moneda, algunas veces, la secuencia de águilas y soles no es tan relevante como el total de águilas o soles. Esas cantidades de interés, o, más formalmente, esas funciones de valores reales definidas en el mismo espacio son conocidas como variables aleatorias.

Como el valor de una variable aleatoria está determinado por el resultado del experimento, es posible asignar probabilidades de los posibles valores de la variable aleatoria.

Ejemplo

Supóngase que un experimento consiste en 3 lanzamientos de una moneda. Si *Y* denota el número de águilas que aparecen, entonces *Y* es una variable aleatoria que toma uno de los valores 0, 1, 2, y 3 con las probabilidades respectivas

$$P\{Y = 0\} = P\{(S, S, S)\} = \frac{1}{8}$$
$$P\{Y = 1\} = P\{(S, S, A), (S, A, S), (A, S, S)\} = \frac{3}{8}$$
$$P\{Y = 2\} = P\{(S, A, A), (A, S, A), (A, A, S)\} = \frac{3}{8}$$
$$P\{Y = 3\} = P\{(A, A, A)\} = \frac{1}{8}$$

Como *Y* debe tomar uno de los valores enteros entre 0 y 3, se debe tener que

$$P\left(\bigcup_{i=0}^{3}\{Y = i\}\right) = \sum_{i=0}^{3} P\{Y = i\} = 1$$

lo cual, de hecho, concuerda con las probabilidades anteriores.

2.1 Variables Aleatorias Discretas

Una variable aleatoria discreta, comúnmente denotada por *X, Y* o similarmente, toma un número finito o infinito contable de posibles valores con probabilidades específicas asociadas con cada valor. En una agencia de cobros, por ejemplo, el administrador puede mirar el patrón del número (*X*) de cuentas morosas. En una planta de empaquetamiento, por ejemplo, alguien puede estar interesado en estudiar el patrón del número (*X*) de paquetes defectuosos. En un camión de carga, por ejemplo, el agente de recepción puede querer estudiar el patrón del número (*X*) de naranjas podridas o el número (*X*) de veces que el camión llega tarde. Alguien puede preguntar: ¿Cuántas veces (*X*) debe alguien perforar en un campo de petróleo para hallar el crudo? Estas son algunos ejemplos típicos de variables aleatorias discretas.

Cuando se lanza una moneda legal, sea *Y* el número de lanzamientos de la moneda requeridos para observar la primera águila (*A*) obtenida. Entonces, $P\{Y=1\}=P\{La\ A\ aparezca\ en\ el\ primer\ lanzamiento\}=P\{A\}=1/2$, y $P\{Y=2\}=P\{La\ primera\ A\ aparezca\ en\ el\ segundo\ lanzamiento\}=P\{(S,A)\}=\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)=\frac{1}{4}$. Similarmente, $P\{Y=3\}=P\{(S,S,A)\}=1/8$, …, que es

$$P\{Y=y\}=\left(\frac{1}{2}\right)^{y}, y=1,2,\ldots \qquad (2.1.1)$$

Así, el conjunto de posibles valores para la variable *Y* es infinito contable. (Mukhopdhyay, 2000)

2.1.1 Funciones de Masa de Probabilidad y Distribución

En general, una variable aleatoria *X* es una asignación (que es, una función) del espacio muestral $\boldsymbol{S}$ a un subconjunto χ de la línea real $\mathbb{R}$, lo cual equivale a decir que la variable *X* induce eventos ($\in\ \beta$) en el contexto de $\boldsymbol{S}$**.** Se puede expresar esto escribiendo $X: S \rightarrow \chi$. En el caso discreto, supóngase que *X* toma los posibles valores $x_1, x_2, x_3, \ldots$ con las respectivas probabilidades $p_i=P\{X=x_i\}$, $i=1,2,\ldots$. Matemáticamente, se evalúa $P\{X=x_i\}$ como sigue:

$$p_i=P\{X=x_i\}=\sum_{s\in S:X(s)=x_i} P(s)\,, \qquad i=1,2,\ldots \qquad (2.1.2)$$

En (2.1.2), se encuentra $P\{X=i\}$ para $i=1,2,\ldots$ es la probabilidad de que en el experimento se obtenga como resultado el valor x_i.

Mientras se asignan o evalúan estas probabilidades, se tiene que estar seguro de que las siguientes dos condiciones se satisfacen:

$$i)\ p_i \geq 0\ para\ toda\ i = 1, 2, \ldots, y$$
$$ii)\ \sum_{i=1}^{\infty} p_i = 1 \qquad (2.1.3)$$

Cuando estas condiciones se cumplen, se llama a una asignación tal como

$$\begin{array}{rl} \textit{Valores de X:} & x_1\ \ x_2\ \ .\ \ .\ \ .\ \ x_i\ \ .\ \ .\ \ . \\ \textit{Probabilidades:} & p_1\ \ p_2\ \ .\ \ .\ \ .\ \ p_i\ \ .\ \ .\ \ . \end{array} \qquad (2.1.4)$$

una distribución discreta o una distribución de probabilidad discreta de la variable *X*.

La función $f(x) = P\{X = x\}\ para\ x \in \chi = \{x_1, x_2, \ldots, x_i, \ldots\}$ es comúnmente conocida como la función de masa de probabilidad (fmp) de *X*.

Además, se define otra útil función asociada con una variable aleatoria *X* como sigue:

$$F(x) = P\{X \leq x\} = P\{s: s \in S\ tal\ que\ X(s) \leq x\}, x \in \mathbb{R} \qquad (2.1.5)$$

la cual es comúnmente llamada la función de distribución (fd) o la función de distribución acumulada (fda) de *X*. Algunas veces, en lugar de $F(x)$ se puede escribir $F_X(x)$ para la fda de la variable aleatoria *X*.

Una vez que la fmp $f(x)$ está dada como en (2.1.4), se puede encontrar las probabilidades de eventos que son definidos a través de la variable aleatoria *X*. Si se denota un conjunto $A(\subseteq \mathbb{R};)$, entonces

$$P\{X \in A\} = \sum_{i: x_i \in A \cap \chi} P\{X = x_i\}. \qquad (2.1.6)$$

(Mukhopdhyay, 2000)

2.2 Variables Aleatorias Continuas

Una variable aleatoria continua, comúnmente denotada por *X*, *Y* o similarmente, toma valores solo dentro de subintervalos de ℝ o dentro de subconjuntos generados por algunos subintervalos apropiados de la línea real ℝ. En alguna ubicación particular en un lago, por ejemplo, el administrador del departamento local de parques y recreación puede estar interesado en estudiar el patrón de profundidad (*X*) del nivel de agua. En un alto edificio de oficinas, por ejemplo, alguien puede estar interesado en investigar el patrón de tiempo de espera (*X*) para un elevador en alguno de sus pisos. En el momento del festival anual de música en un parque de diversiones, por ejemplo, el administrador de un centro de tercera edad cercano puede querer estudiar el patrón del factor de sonoridad (*X*). Éstos son algunos ejemplos típicos de variables aleatorias continuas.

2.2.1 Funciones de Densidad de Probabilidad y de Distribución

Se ha asumido que el espacio muestral $\boldsymbol{S}$ mismo es un subintervalo de $\mathbb{R}$. Ahora, una variable aleatoria continua *X* es una asignación de $\boldsymbol{S}$ en la línea real $\mathbb{R}$. En este escenario, no se puede hablar más de $P\{X = x\}$, ya que esta probabilidad será igual a cero, independientemente de cual valor específico $x \in \mathbb{R}$ se tenga en mente. Por ejemplo, el tiempo de espera (*X*) en una parada de autobús debería estar comúnmente postulado como una variable aleatoria continua, y así la probabilidad de que alguien espere exactamente cinco minutos o exactamente siete minutos en la parada para que llegue el siguiente autobús es simplemente cero.

Para facilitar la modelación de una situación estocástica continua, comiéncese con una función $f(x)$ asociada con cada valor $x \in \mathbb{R}$ satisfaciendo las siguientes dos propiedades:

$$i)\ f(x) \geq 0\ para\ toda\ x \in\ \mathbb{R},\ y$$

$$ii)\ \int_{\mathbb{R}} f(x)dx = 1 \qquad (2.2.1)$$

Obsérvese que las interpretaciones de (2.1.3) y (2.2.1) son realmente muy similares. Las condiciones listadas en (2.2.1) son simplemente los análogos continuos de las listadas en (2.1.3). Una función $f(x)$ que satisface (2.2.1) es llamada una función de densidad de probabilidad (*fdp*).

Una vez que la fdp $f(x)$ es especificada, se pueden encontrar las probabilidades de varios eventos definidos en términos de la variable aleatoria *X*. Si se denota un conjunto $A(\subseteq \mathbb{R})$, entonces

$$P\{X \in A\} = \int_A f(x)dx, \qquad (2.2.2)$$

Donde la convención es que se debería integrar la función $f(x)$ solo en la parte del conjunto A siempre que $f(x)$ es positiva.

En otras palabras, $P\{x \in A\}$ está dada por el área bajo la curva $\{(x, f(x)); para\ todo\ x\ \in A, siempre\ que\ f(x) > 0\}$. En la figura 2.2.1, sea el conjunto A el intervalo (a, b) y el área sombreada representa la probabilidad correspondiente, $P\{a < X < b\}$.

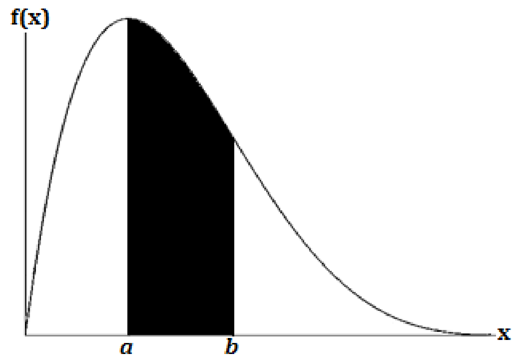

Figura 2.1: El área sombreada bajo la FDP $f(x)$ es $P\{a < X < b\}$. Fuente: Elaboración propia

Se define la función de distribución (fd) de una variable aleatoria continua X modificando la análoga discreta de (2.1.5). Sea

$$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} f(y)dy\,, x \in \mathbb{R} \tag{2.2.3}$$

la cual también es llamada, función de distribución acumulada (fda). Nuevamente, nótese que $F(x)$ es definida por todos los valores reales x. Al igual que en el caso discreto, algunas veces también se puede escribir $F_X(x)$ para la fda de la variable aleatoria X.

2.3 Esperanza y Varianza de una Variable Aleatoria

En esta sección se discutirán los conceptos de esperanza y varianza de una variable aleatoria. El valor esperado o esperanza de una variable aleatoria es algunas veces juzgado como "el centro" de la distribución de probabilidad de la variable. Entonces, la varianza de una variable aleatoria cuantifica la desviación cuadrada promedio de la variable aleatoria desde su "centro".

2.3.1 Esperanza de una Variable Aleatoria

Considérese un juego de azar. La casa lanzará un dado legal. El jugador ganará \$8 de la casa siempre que caiga un seis, pero el jugador pagará \$2 a la casa cuando caiga una cara distinta a seis. Supóngase, por ejemplo, que de 10 turnos sucesivos se observan las siguientes caras en el dado lanzado: 4, 3, 6, 6, 2, 5, 3, 6, 1, 4. En este punto, el jugador se encuentra adelante por \$10 (= \$24 - \$14) así que el promedio de ganancia del jugador por juego es exactamente de \$1. Pero en el largo plazo, ¿Cuál es el valor esperado de la ganancia del jugador por partida? Asúmase que el jugador se encuentra en la partida k, y para ese momento la cara seis aparece n_k veces, mientras que las caras distintas a seis han aparecido $k - n_k$ veces. En este punto, el jugador habrá ganado un monto de $8n_k - 2(k - n_k)$ así que la ganancia del jugador por juego (G_k) debería ser $\frac{8n_k - 2(k-n_k)}{k}$ que se puede reescribir como sigue:

$$G_k = (8)(k^{-1}n_k) + (-1)(1 - (k^{-1}n_k)) \tag{2.3.1}$$

¿Cuál será el valor de G_k cuando k sea muy grande? Interpretando las probabilidades como el límite de las frecuencias relativas se puede decir que $\lim_{k\to\infty} \frac{n_k}{k}$ debería coincidir con la probabilidad de observar la cara seis en un solo lanzamiento de un dado legal, lo cual es nada más que $1/6$. Así, la ganancia final por partida del jugador será:

$$\lim_{k\to\infty} G_k = (8)\left(\frac{1}{6}\right) + (-2)\left(\frac{5}{6}\right) = -1/3 \tag{2.3.2}$$

En otras palabras, en este juego el jugador perderá $0.33 por partida en el largo plazo. Es fácilmente visto de (2.3.2) que multiplicamos los posibles valores de la ganancia por su probabilidad y sumamos esos términos.

Definición 2.3.1 El valor esperado de la variable aleatoria X, denotado por $E(X)$, $E[X]$, o $E\{X\}$, está definido como sigue:

$$\sum_{i:\, x_i \in \chi} x_i f(x_i) \ \ cuando\ X\ es\ discreta, o$$

$$\int_{\chi} x f(x) \ \ cuando\ X\ es\ continua$$

Donde χ es el soporte de X. El valor esperado es también llamado la media de la distribución y frecuentemente se le asigna el símbolo μ.

2.3.2 Varianza de una Variable Aleatoria

Dada una variable aleatoria X con su respectiva función de distribución F, sería extremadamente útil si se fuese capaz de resumir las propiedades esenciales de F por medio de ciertas medidas adecuadamente definidas. Una de tales medidas sería $E[X]$, el valor esperado de X. Sin embargo, aunque $E[X]$ mide el promedio ponderado de los posibles valores de X, no dice nada acerca de la variación, o la dispersión de estos valores. A manera de ejemplo, aunque las variables W, Y, y Z teniendo funciones de masa de probabilidad determinadas por

$$W = 0 \ con\ probabilidad\ 1$$

$$Y = \begin{cases} -1 \ con\ probabilidad\ 1/2 \\ +1 \ con\ probabilidad\ 1/2 \end{cases}$$

$$Z = \begin{cases} -100 \ con\ probabilidad\ 1/2 \\ +100 \ con\ probabilidad\ 1/2 \end{cases} \qquad (2.3.3)$$

todas tienen la misma esperanza (igual a 0), hay una mucho más grande variación en los valores posibles de Y que en los de W (la cual es constante) y en los posibles valores de Z que en los de Y.

Como se espera que X tome valores alrededor de su media $E[X]$, parecería que una forma razonable de medir la posible variación de X sería ver cuán lejos está X de su media, en promedio. Una posible forma de medir esta variación sería considerar la cantidad $E[|X-\mu|]$, donde $\mu = E[X]$. Sin embargo, se vuelve matemáticamente inconveniente el manejar esta cantidad, así que, una cantidad más tratable es usualmente considerada, la esperanza del cuadrado de la diferencia entre X y su media. Así se tiene la siguiente definición.

Definición 2.3.2 Si X es una variable aleatoria con media μ, entonces la varianza de X, denotada por $Var(X)$, $Var[X]$, o $Var\{X\}$, está definida por

$$Var[X] = E[(X-\mu)^2] \qquad (2.3.4)$$

Una fórmula alternativa para $Var[X]$ se deriva como sigue:

$$\begin{aligned} &Var[X] = E[(X-\mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + E[\mu^2] \\ &= E[X^2] - 2\mu\mu + \mu^2 \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned}$$

Así que,

$$Var[X] = E[X^2] - (E[X])^2 \qquad (2.3.5)$$

En otras palabras, la varianza de X es igual al valor esperado de X^2 menos el cuadrado de su valor esperado. En la práctica, esta fórmula frecuentemente ofrece la forma más fácil para calcular $Var[X]$.

Capítulo 3

Algunas Distribuciones de Probabilidad

En este capítulo se listan todas las distribuciones de probabilidad utilizadas durante la elaboración del modelo, así como sus propiedades esenciales, y la importancia de cada una de ellas en el caso tratado.

3.1 Distribución Bernoulli

Ésta es quizá una de las más simples variables aleatorias discretas. Se dice que una variable aleatoria X tiene la distribución Bernoulli (p) si y sólo si su fmp está dada por:

$$f(x) = P\{X = x\} = p^x(1-p)^{1-x} \; para \; x = 0, 1 \tag{3.1.1}$$

donde $0 < p < 1$. Aquí, p es a menudo referida como un parámetro. En las aplicaciones, se pueden coleccionar datos dicótomos, por ejemplo, simplemente registrar si un artículo está defectuoso $(x = 0)$ o no defectuoso $(x = 1)$, si un individuo está casado $(x = 0)$ o no lo está $(x = 1)$, o si una vacuna funciona $(x = 1)$ o no funciona $(x = 1)$, y similarmente. En cada situación, p representa $P\{X = 1\}$ y $1 - p$ representa $P\{X = 0\}$.

De acuerdo con la Definición 2.3.1 y la Definición 2.3.2 y utilizando (2.3.5), se tiene

$$\mu = E[X] = (1)(p) + (0)(1-p) = p, \tag{3.1.2}$$
$$E[X^2] = (1)^2(p) + (0)^2(1-p) = p, \tag{3.1.3}$$
$$\sigma^2 = Var[X] = E[X^2] - (E[X])^2 = p - p^2 = p(1-p) \tag{3.1.4}$$

3.2 Distribución Uniforme

Una variable aleatoria continua X tiene distribución uniforme con parámetros a y b, denotado por $U(a, b)$ o $Uniforme(a, b)$, si y sólo si, su función de densidad se define por:

$$f_X(x) = \frac{1}{b-a} \; para \; a < x < b \tag{3.2.1}$$

donde $-\infty < a, b < \infty$.

Dado que la función de densidad de una variable aleatoria uniforme es una constante $\left(\frac{1}{b-a}\right)$, el área comprendida entre la gráfica de dicha función, y el segmento (a, b) del eje x, corresponde a un rectángulo, cuyas dimensiones son $b - a, y \frac{1}{b-a}$, este hecho facilita el cálculo de la función de distribución. Además, como la función de distribución acumulada $F_X(x)$ es el área bajo la curva de la función de densidad, definida desde $-\infty$ hasta x, esta área corresponde nuevamente a un rectángulo, esta vez de dimensiones $x - a, y \frac{1}{b-a}$. Por lo que, la función de distribución acumulada para una variable aleatoria uniforme es la siguiente:

$$F_X(x) = \frac{x-a}{b-a}, a < x < b \qquad (3.2.2)$$

Ahora, utilizando la definición (2.3.1) y (3.2.1), la esperanza de una variable aleatoria uniforme está definida como sigue:

$$\begin{aligned} E[X] &= \int_a^b \frac{x}{b-a} dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b^2-a^2}{2(b-a)} \\ &= \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2} \end{aligned} \qquad (3.2.3)$$

Que, de hecho, es el punto medio entre $a\ y\ b$. Para obtener la varianza, se hará uso de la ecuación (2.3.5).

$$\begin{aligned} E[X^2] &= \int_a^b \frac{x^2}{b-a} dx = \left.\frac{x^3}{3(b-a)}\right|_a^b = \frac{b^3-a^3}{3(b-a)} \\ &= \frac{(b^2+ab+a^2)(b-a)}{3(b-a)} = \frac{b^2+ab+a^2}{3} \\ \Rightarrow Var[X] &= \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{4b^2+4ab+4a^2-3(a^2+2ab+b^2)}{12} \\ &= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12} \end{aligned} \qquad (3.2.4)$$

La distribución uniforme, es la distribución más simple para modelar variables acotadas, ya que cada valor – desde el mínimo, hasta el máximo – son igualmente probables.

La distribución uniforme es útil en situaciones en las cuales se tienen un máximo y un mínimo disponibles, pero ninguna otra información. Es fácil ver que la mayoría de las situaciones del mundo real no caen en este modelo, y en muchos casos es posible tener una estimación adicional de un valor más probable, por ejemplo. En caso de tener una estimación de dicho valor, es posible utilizarla para construir un modelo más realista, como los que se presentan a continuación.

3.3 Distribución Triangular

Si se tiene una estimación de un “valor más probable” además del mínimo y el máximo, se puede utilizar esta información adicional para construir una distribución de probabilidad que favorezca el modelo. La distribución más simple en este contexto es la distribución triangular, la gráfica de la densidad de esta distribución se asemeja a un triángulo con el valor más probable (llamado moda) en el vértice superior del triángulo.

A diferencia de una distribución uniforme, la distribución triangular enfatiza el valor más probable, lo cual, teóricamente, debería proporcionar una mejor estimación de las probabilidades de obtener otros valores.

Si se asume que a es el valor mínimo que puede tomar la variable; b el valor máximo de la variable; y m el valor más probable, o moda, como el nombre lo indica, el área comprendida entre la curva de la función de densidad, y el segmento (a, b) del eje x, corresponde a un triángulo, cuyos vértices son los puntos $(a, 0), (m, h)$ y $(b, 0)$, donde h representa la altura del triángulo.

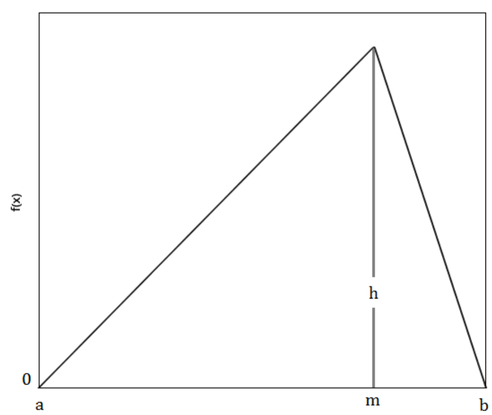

Figura 3.1: Función de densidad de la distribución triangular. Fuente: Elaboración propia

Como se puede observar en la figura 3.2, la función de densidad de una variable aleatoria triangular cambia su comportamiento a partir de la moda, por lo que dicha función está dividida en casos, y si se toma en cuenta que el área bajo la curva de densidad de cualquier variable aleatoria continua es igual a 1, utilizando la fórmula para el área de un triángulo se puede deducir lo siguiente:

$$\frac{h(b-a)}{2} = 1 \qquad\qquad \Rightarrow h = \frac{2}{b-a}$$

De esa forma, utilizando la fórmula para la ecuación de una recta:

$$y = p(x - x_1) + y_1$$

$$p = \frac{y_2 - y_1}{x_2 - x_1}, y$$

$$(x_1, y_1) = (a, 0)$$

$$(x_2, y_2) = \left(m, \frac{2}{b - a}\right)$$

$$(x_3, y_3) = (b, 0)$$

$$\Rightarrow f_1(x) = \frac{2(x - a)}{(b - a)(m - a)},$$

$$f_2(x) = \frac{2(b - x)}{(b - a)(b - m)}$$

$$\therefore f_X(x) = \begin{cases} \dfrac{2(x - a)}{(b - a)(m - a)}, & si\ a \leq x < m \\ \dfrac{2}{b - a}, & si\ x = m \\ \dfrac{2(b - x)}{(b - a)(b - m)}, & si\ m < x \leq b \end{cases} \qquad (3.3.1)$$

De la misma forma en que la función de densidad se comporta distintamente antes y después de la moda, la función de distribución acumulada tiene un comportamiento similar. Para calcular la función de distribución acumulada, es necesario notar que para los valores x que cumplen $a < x \leq m$, el área que está bajo la función de densidad y sobre el segmento $(a, x]$ corresponde nuevamente a un triángulo, cuya base es igual a $x - a$ y su altura es igual a $f_X(x) = \frac{2(x-a)}{(b-a)(m-a)}$. Mientras que, si $m < x < b$, el área comprendida bajo la función de densidad y sobre el segmento (x, b) corresponde a un triángulo con base igual a $b - x$ y una altura igual a $f_X(x) = \frac{2(b-x)}{(b-a)(b-m)}$. Tomando en cuenta lo anterior, y utilizando la fórmula para calcular el área de un triángulo, la función de distribución acumulada para una variable aleatoria triangular está dada por:

$$F_x(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(m-a)} \times \left(\dfrac{x-a}{2}\right), si\ a < x \leq m \\ 1 - \dfrac{2(b-x)}{(b-a)(b-m)} \times \left(\dfrac{b-x}{2}\right), si\ m < x < b \end{cases}$$

$$\Rightarrow F_x(x) = \begin{cases} \dfrac{(x-a)^2}{(b-a)(m-a)}, si\ a < x \leq m \\ 1 - \dfrac{(b-x)^2}{(b-a)(b-m)}, si\ m < x < b \end{cases} \tag{3.3.2}$$

Utilizando la definición 2.3.1 y la ecuación (3.3.1, se calcula la esperanza:

$$\begin{aligned} E[X] &= \int_a^m \frac{2x(x-a)}{(b-a)(m-a)} dx + \int_m^b \frac{2x(b-x)}{(b-a)(b-m)} dx \\ &= \frac{(m-a)(a+2m)}{3(b-a)} + \frac{(b-m)(b+2m)}{3(b-a)} \\ &= \frac{am + 2m^2 - a^2 - 2am + b^2 + 2bm - bm - 2m^2}{3(b-a)} \\ &= \frac{b^2 + bm - am - a^2}{3(b-a)} = \frac{(b-a)[(b+a)+m]}{3(b-a)} \\ &= \frac{a+m+b}{3} \end{aligned} \tag{3.3.3}$$

Para calcular $Var[X]$ se utilizará (2.3.5) y (3.3.1)

$$\begin{aligned} E[X^2] &= \int_a^m \frac{2x^2(x-a)}{(b-a)(m-a)} dx + \int_m^b \frac{2x^2(b-x)}{(b-a)(b-m)} dx \\ &= \frac{3m^3 - a^2m - am^2 - a^3}{6(b-a)} + \frac{(b-m)(2bm + b^2 + 3m^2)}{6(b-a)} \\ &= \frac{3m^3 - a^2m - am^2 - a^3 + b^3 + b^2m + bm^2 - 3m^3}{6(b-a)} \\ &= \frac{b^3 + (b-a)m^2 + (b^2-a^2)m - a^3}{6(b-a)} \\ &= \frac{(b-a)[(b^2+ab+a^2) + m^2 + (b+a)m]}{6(b-a)} \end{aligned}$$

$$= \frac{(b^2 + ab + a^2) + m^2 + (b + a)m}{6}$$

$$\Rightarrow Var[X] = \frac{(b^2 + ab + a^2) + m^2 + (b + a)m}{6} - \left(\frac{a + m + b}{3}\right)^2$$

$$= \frac{a^2 + m^2 + b^2 - am - ab - bm}{18} \qquad (3.3.4)$$

3.4 Distribución Beta

Una variable aleatoria continua X tiene distribución $Beta(\alpha, \beta)$ si y sólo si, su función de densidad está dada por:

$$f_X(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}, & 0 \le x \le 1 \\ 0, & en\ otro\ caso \end{cases} \qquad (3.4.1)$$

donde los parámetros $\alpha > 0, \beta > 0$ y $\mathrm{B}(\alpha, \beta)$ es la función Beta que está definida de la siguiente manera:

$$\mathrm{B}(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \qquad (3.4.2)$$

donde $\Gamma(\alpha), \Gamma(\beta), y\ \Gamma(\alpha + \beta)$, representan la función gamma ($\Gamma(z)$), evaluada en α, β, y $\alpha + \beta$, correspondientemente, y dicha función equivale a la siguiente integral:

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt \qquad (3.4.3)$$

La función de distribución acumulada para una variable aleatoria Beta con parámetros $\alpha\ y\ \beta$ se muestra en la siguiente ecuación

$$F_X(x) = \frac{\mathrm{B}_x(\alpha, \beta)}{\mathrm{B}(\alpha, \beta)} \qquad (3.4.4)$$

donde

$$\mathrm{B}_x(\alpha, \beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt \qquad (3.4.5)$$

A la función anterior se le conoce como función beta incompleta en x, y cuando $x = 1$, esta función coincide con la función beta completa.

Según la definición (2.3.1) y la ecuación (3.2.2), la varianza de una variable aleatoria con distribución beta es

$$E[X] = \int_0^1 x \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} dx = \frac{1}{\mathrm{B}(\alpha,\beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1} dx$$

$$= \frac{\mathrm{B}(\alpha+1,\beta)}{\mathrm{B}(\alpha,\beta)} = \frac{\Gamma(\alpha+1)\Gamma(\beta)/\Gamma(\alpha+\beta+1)}{\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)} = \frac{\Gamma(\alpha+1)\Gamma(\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+1)}$$

$$= \frac{(\alpha)\Gamma(\alpha)\Gamma(\alpha+\beta)}{\Gamma(\alpha)((\alpha+\beta)\Gamma(\alpha+\beta))} = \frac{\alpha}{\alpha+\beta} \tag{3.4.6}$$

De igual forma,

$$E[X^2] = \int_0^1 x^2 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} dx = \frac{1}{\mathrm{B}(\alpha,\beta)} \int_0^1 x^{\alpha+1}(1-x)^{\beta-1} dx$$

$$= \frac{\mathrm{B}(\alpha+2,\beta)}{\mathrm{B}(\alpha,\beta)} = \frac{\Gamma(\alpha+2)\Gamma(\beta)/\Gamma(\alpha+\beta+2)}{\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)} = \frac{\Gamma(\alpha+2)\Gamma(\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+2)}$$

$$= \frac{\alpha(\alpha+1)\Gamma(\alpha)\Gamma(\alpha+\beta)}{\Gamma(\alpha)(\alpha+\beta)(\alpha+\beta+1)\Gamma(\alpha+\beta)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \tag{3.4.7}$$

Sustituyendo (3.2.6) y (3.2.7) en (2.3.5), la varianza de una variable aleatoria X con distribución beta se define por:

$$Var[X] = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2$$

$$= \frac{\alpha(\alpha+1)(\alpha+\beta) - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{3.4.8}$$

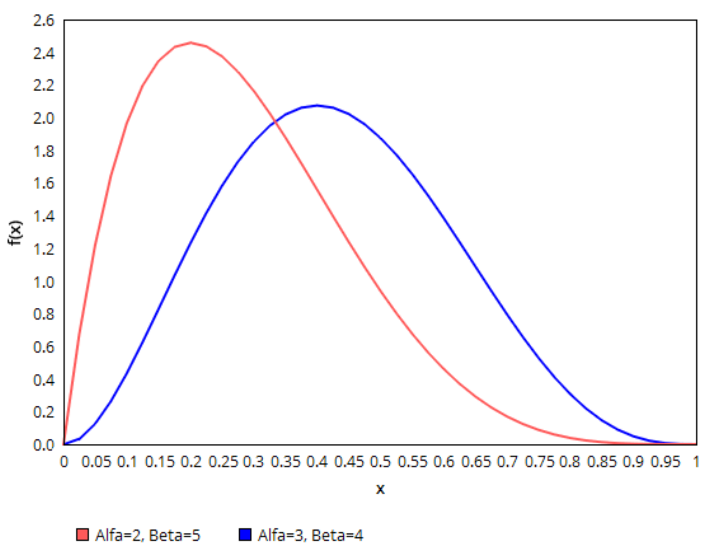

Figura 3.2: Función de densidad de la distribución beta. Fuente: Elaboración propia

La distribución beta puede ser usada para modelar un fenómeno aleatorio cuyo conjunto de valores posibles es algún intervalo finito $[a, b]$, de tal forma que si a representa el origen y se toma $b - a$ como una medida unitaria, puede ser transformada en el intervalo $[0,1]$.

Dado que el intervalo en que está definida la distribución beta, esta distribución se utiliza frecuentemente para modelar proporciones, por ejemplo, la proporción del tiempo de vida de algún aparato, durante el cual se encuentra en reparación.

3.5 Distribución PERT

La distribución PERT o distribución beta-PERT es una útil herramienta para modelar datos expertos. Cuando es usada en una simulación Monte Carlo, la distribución PERT puede ser usada para identificar riesgos en proyectos y modelos de costos basados en la probabilidad del cumplimiento de objetivos y metas por medio de cualquier número de componentes del proyecto.

De la misma forma que con cualquier distribución de probabilidad, la utilidad de la distribución PERT está limitada por la calidad de las estimaciones: cuanto mejor sea la estimación, se obtendrán mejores resultados en una simulación.

En planeaciones de costos o proyectos, se pueden poseer estimaciones del tiempo esperado, costos u otras variables. Para construir modelos probabilísticos, es necesario un rango de estimaciones – un valor mínimo y un valor máximo, y si es posible, un valor más probable. Estas estimaciones son usadas para construir una distribución de probabilidad, y entonces una simulación Monte Carlo puede ser corrida basada en esa distribución.

Cuando se tiene un rango de estimación disponible, hay varias formas diferentes con las que se puede modelar la distribución o generar valores muestrales. La distribución uniforme, la distribución triangular, y la distribución PERT son tres de ellas.

La distribución PERT es un caso especial de la distribución beta que utiliza 3 parámetros: un mínimo (a), un máximo (b) y un valor más probable (m), a este último también se le conoce como moda. A diferencia de la distribución triangular, la distribución PERT usa estos parámetros para crear una curva suave que se asimila a la distribución normal o la lognormal. (The beta-PERT distribution, s.f.)

Supóngase que se conoce el mínimo (a), el máximo (b) y la moda (m) de una variable aleatoria, y sea X una variable aleatoria con distribución Beta y parámetros $\alpha\ y\ \beta$, donde

$$\alpha = \frac{4m + b - 5a}{b - a} \tag{3.5.1}$$

$$\beta = \frac{5b - a - 4m}{b - a} \tag{3.5.2}$$

Si $Y = (b - a)X + a$, entonces Y tiene una distribución PERT con parámetros $a, b\ y\ m$. Sabiendo lo anterior, es posible construir la función de densidad para una variable aleatoria con distribución PERT, como se muestra a continuación:

$$Y = g(X) = (b - a)X + a \tag{3.5.3}$$

$$F_Y(y) = P\{Y \le y\} = P\{g(X) \le y\}$$

$$= P\{X \le g^{-1}(y)\} = P\left\{X \le \frac{y-a}{b-a}\right\} = F_X\left(\frac{y-a}{b-a}\right) \tag{3.5.4}$$

$$\Rightarrow f_Y(y) = f_X\left(\frac{y-a}{b-a}\right) \times \frac{dg^{-1}(y)}{dy} = f_X\left(\frac{y-a}{b-a}\right)\left(\frac{1}{b-a}\right)$$

Sustituyendo en (3.4.1)

$$\Rightarrow f_Y(y) = \frac{\left(\frac{y-a}{b-a}\right)^{4\left(\frac{m-a}{b-a}\right)}\left(\frac{b-y}{b-a}\right)^{4\left(\frac{b-m}{b-a}\right)}}{(b-a)\mathrm{B}\left(\frac{4m+b-5a}{b-a}, \frac{5b-a-4m}{b-a}\right)}$$

$$\text{, para } a \le x \le b$$

$$\therefore f_Y(y) = \frac{5!\,(y-a)^{4\left(\frac{m-a}{b-a}\right)}(b-y)^{4\left(\frac{b-m}{b-a}\right)}}{(b-a)^5\Gamma\left(\frac{4m+b-5a}{b-a}\right)\Gamma\left(\frac{5b-a-4m}{b-a}\right)} \tag{3.5.5}$$

Haciendo uso de (3.4.4) y (3.5.4) se obtiene la siguiente fda:

$$F_Y(y) = \frac{\mathrm{B}_{\frac{y-a}{b-a}}\left(\frac{4m+b-5a}{b-a}, \frac{5b-a-4m}{b-a}\right)}{\mathrm{B}\left(\frac{4m+b-5a}{b-a}, \frac{5b-a-4m}{b-a}\right)}$$

$$\text{, para } a \le x \le b \tag{3.5.6}$$

Para calcular la esperanza de una variable aleatoria con distribución PERT, basta con sustituir (3.5.1) y (3.5.2) en (3.4.6) y hacer uso de (3.5.3)

$$E[Y] = E[(b-a)X] + a$$

$$= (b-a)E[X] + a = (b-a)\frac{\left(\frac{4m+b-5a}{b-a}\right)}{\left(\frac{4m+b-5a}{b-a}+\frac{5b-a-4m}{b-a}\right)} + a$$

$$= (b-a)\left(\frac{4m+b-5a}{6b-6a}\right) + a = \frac{4m+b-5a+6a}{6}$$
$$= \frac{a+4m+b}{6}$$
$$\therefore E[Y] = \frac{a+4m+b}{6} \tag{3.5.7}$$

Utilizando, el mismo procedimiento, ahora para calcular la varianza, sustituyendo (3.5.1) y (3.5.1) en (3.4.8) y (3.5.3)

$$Var[Y] = Var[(b-a)X + a] = (b-a)^2 Var[X]$$

$$= \frac{(b-a)^2\left(\frac{4m+b-5a}{b-a}\right)\left(\frac{5b-a-4m}{b-a}\right)}{\left(\frac{4m+b-5a+5b-a-4m}{b-a}\right)^2 + \left(\frac{4m+b-5a+5b-a-4m}{b-a}+1\right)}$$

$$= \frac{(4m+b-5a)(5b-a-4m)}{(6)^2(6+1)} = \frac{(4m+b-5a)(5b-a-4m)}{252}$$

$$\therefore Var[Y] = \frac{(4m+b-5a)(5b-a-4m)}{252} \tag{3.5.8}$$

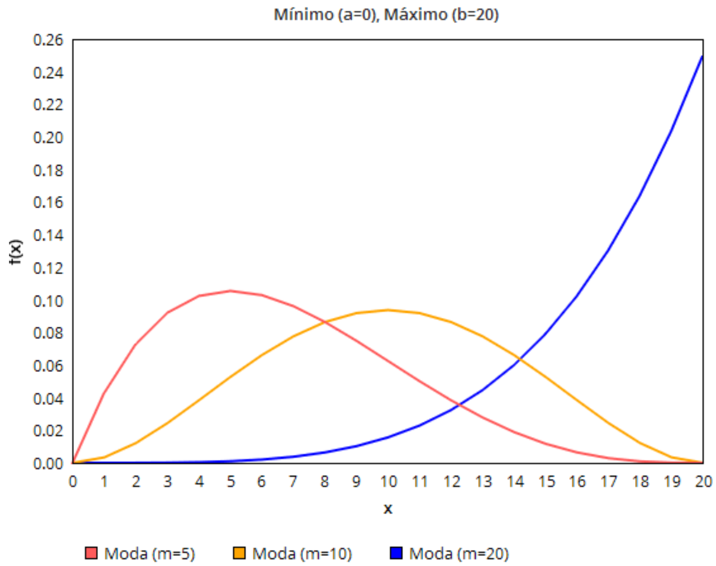

Figura 3.3: Función de densidad de la distribución PERT. Fuente: Elaboración propia

3.6 Comparación entre las distribuciones Uniforme, Triangular y PERT

De acuerdo con lo anterior, la distribución uniforme – para la opinión experta – es un modelo muy pobre, dado que cada uno de los valores abarcados en su rango tienen una densidad de probabilidad igual, pero esa densidad cae fuertemente a cero en el valor mínimo y en el máximo de una sobrenatural manera. Se sabe que la distribución uniforme obedece el hecho de que sólo se conocen el mínimo y el máximo, pero es raro que un experto sólo pueda definir estos 2 parámetros sin poder hacerlo con un valor más probable. Sin embargo, la distribución uniforme puede ser utilizada para resaltar o exagerar el hecho de que se conoce poco acerca de dicho parámetro. (Uniform distribution, s.f.)

Supóngase que se desea modelar un fenómeno aleatorio cuyos valores mínimo y máximo son $a\ y\ b$ respectivamente, de esta forma, si la variable se modela de acuerdo con una distribución uniforme, la varianza del fenómeno estará dada por la ecuación (3.2.4), es decir, $(b-a)^2/12$, por la densidad de la distribución uniforme puede resultar obvio, pero es importante resaltar que la varianza sólo depende de los parámetros $a\ y\ b$ y de ningún otro, razón por la cual nunca variará si se utilizan los mismos parámetros.

Por otro lado, si se contara con información adicional del fenómeno, por ejemplo, el valor más probable m, podría construirse un modelo por medio de la distribución triangular, o bien, de la distribución PERT, por lo tanto, la varianza del fenómeno estaría representada por (3.3.5) o (3.5.8), es decir, $\frac{a^2+m^2+b^2-am-ab-bm}{18}$ o $\frac{(4m+b-5a)(5b-a-4m)}{252}$ que corresponden a las varianzas de las distribuciones triangular y PERT, respectivamente. A continuación, se muestra un contraste entre las varianzas de las tres distribuciones, utilizando técnicas del cálculo diferencial para la determinación de valores críticos.

Como la varianza de la distribución uniforme sólo depende de los parámetros $a\ y\ b$, no se le dará el mismo tratamiento que a las demás distribuciones. Ahora, defínanse las siguientes funciones de la moda m:

$$\varphi_1(m) = \frac{a^2+m^2+b^2-am-ab-bm}{18}$$

$$\varphi_2(m) = \frac{(4m+b-5a)(5b-a-4m)}{252}$$

Comenzando con la distribución triangular, si se deriva $\varphi_1(m)$ se obtiene lo siguiente:

$$\varphi_1'(m) = \frac{2m - (a + b)}{18}$$

Utilizando el criterio de la derivada para la determinación de valores críticos, se iguala $\varphi'_1(m)$ a cero, y se obtiene el valor crítico de la siguiente manera:

$$\frac{2m - (a + b)}{18} = 0 \quad \Rightarrow \quad \frac{2m}{18} = \frac{a + b}{18} \qquad \Rightarrow m = \frac{a + b}{2}$$

Con esto, se puede afirmar que el valor $(a + b)/2$ corresponde a un valor crítico en la varianza de la distribución triangular, sin embargo, aún queda determinar si se trata de un mínimo o un máximo, para ello se utiliza el criterio de la segunda derivada como se muestra en seguida.

$$\varphi_1''(m) = \frac{2}{18} > 0 \quad \Rightarrow \quad \varphi_1''\left(\frac{a + b}{2}\right) > 0$$

Como la segunda derivada de $\varphi_1(m)$ es mayor que cero cuando $m = (a + b)/2$, la varianza de la distribución triangular posee un mínimo en $(a + b)/2$. Además:

$$\varphi_1\left(\frac{a + b}{2}\right) = \frac{a^2 + \left(\frac{a + b}{2}\right)^2 + b^2 - (a + b)\left(\frac{a + b}{2}\right) - ab}{18}$$

$$= \frac{4a^2 + (a + b)^2 + 4b^2 - 2(a + b)^2 - 4ab}{4 \times 18}$$

$$= \frac{4a^2 + 4b^2 - a^2 - b^2 - 2ab - 4ab}{72} = \frac{3a^2 + 3b^2 - 6ab}{72}$$

$$= \frac{3(a^2 + b^2 - 2ab)}{3 \times 24} = \frac{(b - a)^2}{24} \qquad (3.6.1)$$

Por lo tanto, la varianza de la distribución triangular alcanza su valor mínimo en el punto $\left(\frac{a+b}{2}, \frac{(b-a)^2}{24}\right)$.

Es importante mencionar que cuanto mayor sea la distancia de un valor en el intervalo $[a, b]$ al valor $(a + b)/2$ mayor será la varianza, la prueba de ello se muestra a continuación.

Sean $X_1 \, y \, X_2$ variables aleatorias con distribución triangular y parámetros (a, b, m_1) y (a, b, m_2) respectivamente, y son tales que:

$$\left|m_1 - \frac{a + b}{2}\right| > \left|m_2 - \frac{a + b}{2}\right|$$

$$\Rightarrow \left(m_1 - \frac{a + b}{2}\right)^2 > \left(m_2 - \frac{a + b}{2}\right)^2$$

$$\Rightarrow \quad \frac{[2m_1 - (a+b)]^2}{4} > \frac{[2m_2 - (a+b)]^2}{4}$$

$$\Rightarrow \quad \frac{4m_1^2 - 4m_1(a+b) + (a+b)^2}{4} > \frac{4m_2^2 - 4m_2(a+b) + (a+b)^2}{4}$$

$$\Rightarrow \quad m_1^2 - m_1(a+b) > m_2^2 - m_2(a+b)$$

$$\Rightarrow \; m_1^2 - m_1(a+b) + a^2 + \; b^2 - ab > m_2^2 - m_2(a+b) + a^2 + \; b^2 - ab$$

$$\Rightarrow \quad \frac{a^2 + \; b^2 + m_1^2 - (a+b)m_1 - ab}{18} > \frac{a^2 + \; b^2 + m_2^2 - (a+b)m_2 - ab}{18}$$

Por (3.2.4) se sabe que los términos de la izquierda y derecha de la desigualdad anterior, corresponden a las varianzas de las variables $X_1 \, y \, X_2$ definidas anteriormente. Por lo tanto, cuanto mayor es la distancia de una moda m en el intervalo $[a.b]$ al valor $(a+b)/2$, mayor es la varianza de la variable asociada.

Es fácil notar que en un intervalo $[a.b]$, los valores más alejados de $(a+b)/2$ son $a \; y \; b$, con esta observación, no se halla problema alguno al afirmar que la varianza de la distribución uniforme alcanzo sus valores máximos cuando $m = a$ y cuando $m = b$. Además, $\varphi_1(a) = \varphi_1(b)$, esto se prueba en seguida:

$$\varphi_1(a) = \frac{a^2 + b^2 + a^2 - ab - ab - a^2}{18} = \frac{a^2 - 2ab + b^2}{18} = \frac{(b-a)^2}{18}$$

$$\varphi_1(b) = \frac{a^2 + b^2 + b^2 - ab - ab - b^2}{18} = \frac{a^2 - 2ab + b^2}{18} = \frac{(b-a)^2}{18}$$

$$\therefore \varphi_1(a) = \frac{(b-a)^2}{18} = \varphi_1(b)$$

Dado lo anterior, se concluye que la distribución triangular alcanza su valor mínimo en el punto $\left(\frac{a+b}{2}, \frac{(b-a)^2}{24}\right)$, y su máximo en los puntos $\left(a, \frac{(b-a)^2}{18}\right) \; y \; \left(b, \frac{(b-a)^2}{18}\right)$.

De la misma forma que con la distribución triangular, en el resto de esta sección se muestra un análisis para la distribución PERT. Anteriormente, se definió la función $\varphi_2(m)$ que no es otra cosa que la varianza de la distribución PERT vista como una función del parámetro m. Para facilitar los cálculos posteriores bajo este párrafo, se muestra la función $\varphi_2(m)$ escrita en una forma más desarrollada.

$$\varphi_2(m) = \frac{5a^2 - 26ab + 16am + 5b^2 + 16bm - 16m^2}{252}$$

$$\Rightarrow \quad \varphi_2'(m) = \frac{16(a+b-2m)}{252}$$

Utilizando los criterios de la primera y segunda derivada se prosigue de la siguiente forma:

$$\frac{16(a+b-2m)}{252} = 0 \quad \Rightarrow \quad a+b-2m = 0 \quad \Rightarrow \quad m = \frac{a+b}{2}$$

$$\Rightarrow \quad \varphi_2''(m) = -\frac{32}{252} = -\frac{8}{63} < 0 \quad \Rightarrow \quad \varphi_2''\left(\frac{a+b}{2}\right) < 0$$

Por consiguiente, la distribución PERT presenta un máximo en su varianza cuando $m = (a+b)/2$.

Al contrario que en la distribución triangular, en la distribución PERT sucede que cuanto más alejado se encuentre el valor de la moda m de $(a+b)/2$, la varianza será menor, esto resulta lógico, ya que en la distribución triangular el valor crítico $(a+b)/2$ representa un mínimo, mientras que en la distribución PERT un máximo.

Para probar lo anterior, defínanse $Y_1\ y\ Y_2$ como variables aleatorias con distribución PERT cuyos parámetros son $(a, b, m_1)\ y\ (a, b, m_2)$ y los valores $m_1\ y\ m_2$ poseen la siguiente característica:

$$\left|m_1 - \frac{a+b}{2}\right| > \left|m_2 - \frac{a+b}{2}\right|$$

$$\Rightarrow \quad \left(m_1 - \frac{a+b}{2}\right)^2 > \left(m_2 - \frac{a+b}{2}\right)^2$$

$$\Rightarrow \quad \frac{[2m_1 - (a+b)]^2}{4} > \frac{[2m_2 - (a+b)]^2}{4}$$

$$\Rightarrow \quad \frac{4m_1^2 - 4m_1(a+b) + (a+b)^2}{4} > \frac{4m_2^2 - 4m_2(a+b) + (a+b)^2}{4}$$

$$\Rightarrow \quad m_1^2 - m_1(a+b) > m_2^2 - m_2(a+b)$$

$$\Rightarrow \quad -16[m_1^2 - m_1(a+b)] < -16[m_2^2 - m_2(a+b)]$$

$$\Rightarrow \quad -16[m_1^2 - m_1(a+b)] + 5a^2 - 26ab + 5b^2$$

$$< -16[m_2^2 - m_2(a+b)] + 5a^2 - 26ab + 5b^2$$

$$\Rightarrow \quad \frac{5a^2 - 26ab + 16am_1 + 5b^2 + 16bm_1 - 16m_1^2}{252}$$

$$< \frac{5a^2 - 26ab + 16am_2 + 5b^2 + 16bm_2 - 16m_2^2}{252}$$

$$\Rightarrow \quad \frac{(4m_1 + b - 5a)(5b - a - 4m_1)}{252} < \frac{(4m_2 + b - 5a)(5b - a - 4m_2)}{252}$$

En la desigualdad anterior, de acuerdo con (3.5.8), los términos de los lados izquierdo y derecho corresponden a las varianzas de las variables $Y_1 \, y \, Y_2$ respectivamente.

∴ Si $Y_1 \sim PERT(a, b, m_1)$, $Y_2 \sim PERT(a, b, m_2)$, y los valores de los parámetros m_1 y m_2 son tales que $\left|m_1 - \frac{a+b}{2}\right| > \left|m_2 - \frac{a+b}{2}\right|$, entonces $Var[Y_1] < Var[Y_2]$.

Como se mencionó anteriormente, los valores más distantes del valor $(a + b)/2$ en un intervalo de la forma $[a, b]$ son a y b, utilizando la afirmación del párrafo anterior, se concluye que los valores donde la distribución PERT presenta sus valores mínimos en a y b.

Ahora que se tiene conocimiento de los valores máximos y mínimos de la varianza de una distribución PERT, resta evaluar la función $\varphi_2(m)$ en dichos valores. Al igual que en la distribución triangular, en la distribución PERT se tiene que $\varphi_2(a) = \varphi_2(b)$, es decir, el valor de la varianza de dos variables aleatorias triangulares o PERT, es igual cuando una de ellas tiene una moda igual al mínimo que cuando la otra tiene una moda igual al máximo. Lo anterior, se prueba a continuación:

$$\varphi_2(a) = \frac{(4a + b - 5a)(5b - a - 4a)}{252} = \frac{(b - a)(5b - 5a)}{252} = \frac{5(b - a)^2}{252}$$

$$\varphi_2(b) = \frac{(4b + b - 5a)(5b - a - 4b)}{252} = \frac{(5b - 5a)(b - a)}{252} = \frac{5(b - a)^2}{252}$$

$$\therefore \varphi_2(a) = \frac{5(b - a)^2}{252} = \frac{(b - a)^2}{50.4} = \varphi_2(b)$$

Ahora, el valor de la función $\varphi_2(m)$ evaluada en $(a + b)/2$ está dado por:

$$\varphi_2\left(\frac{a + b}{2}\right) = \frac{[2(a + b) + b - 5a][5b - a - 2(a + b)]}{252} = \frac{(3b - 3a)^2}{252}$$

$$= \frac{9(b - a)^2}{252} = \frac{(b - a)^2}{28}$$

$$\therefore \varphi_2\left(\frac{a + b}{2}\right) = \frac{(b - a)^2}{28}$$

Por lo tanto, la distribución PERT con parámetros mínimo y máximo, a y b, respectivamente, alcanza el valor máximo de su varianza cuando la moda es igual a $(a+b)/2$ y el valor de la varianza es $(b-a)^2/28$, mientras que presenta su valor mínimo cuando la moda es igual a a o b, y el valor de la varianza es igual a $(b-a)^2/50.4$.

A continuación, se muestra una tabla que muestra un resumen de los análisis presentados en esta sección.

DISTRIBUCIÓN	*Valor de la moda en el cual la varianza alcanza el mínimo.*	*Valor mínimo de la varianza.*	*Valor de la moda en el cual la varianza alcanza el máximo.*	*Valor máximo de la varianza.*
Uniforme	$\sin moda$	$\frac{(b-a)^2}{12}$	$\sin moda$	$\frac{(b-a)^2}{12}$
Triangular	$\frac{a+b}{2}$	$\frac{(b-a)^2}{24}$	$a\ y\ b$	$\frac{(b-a)^2}{18}$
PERT	$a\ y\ b$	$\frac{(b-a)^2}{50.4}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{28}$

Tabla 3.1: Comparación entre las varianzas de las distribuciones Uniforme, Triangular y PERT. Fuente: Elaboración propia

Nótese en la tabla anterior que, puesto que $(b-a)^2$ es un valor positivo, entonces $(b-a)^2/18 < (b-a)^2/12$, es decir, el máximo valor que puede tomar la varianza de una variable aleatoria con distribución triangular cuyos valores mínimo y máximo son a y b respectivamente, es siempre menor al valor que toma la varianza de una variable aleatoria con distribución uniforme, cuyos parámetros son iguales a los de dicha variable triangular. Además, $(b-a)^2/28 < (b-a)^2/24$, es decir, el máximo valor que puede tomar la varianza de una variable aleatoria con distribución PERT y cutos parámetros mínimo y máximo son respectivamente a y b es menor al mínimo valor que puede tomar una variable aleatoria con distribución triangular y los mismos parámetros.

Las aseveraciones del párrafo anterior significan que siempre que se tenga una variable aleatoria con distribución uniforme, una con distribución triangular, y otra con distribución PERT (todas ellas con mínimos y máximos iguales a a y b); la varianza de la variable PERT siempre será menor a la de la variable triangular, y la varianza de la variable triangular será siempre menor a la de la variable uniforme, y en consecuencia, la variable PERT tendrá una varianza menor a la de la variable triangular, esto sin importar cuales sean las modas de las variables triangular y PERT. Si se interpretan adecuadamente, estas afirmaciones pueden fungir como una alternativa para modelar un fenómeno aleatorio cuando no se tiene información precisa acerca de la moda, o suficientes observaciones para utilizar un método de ajuste más sofisticado.

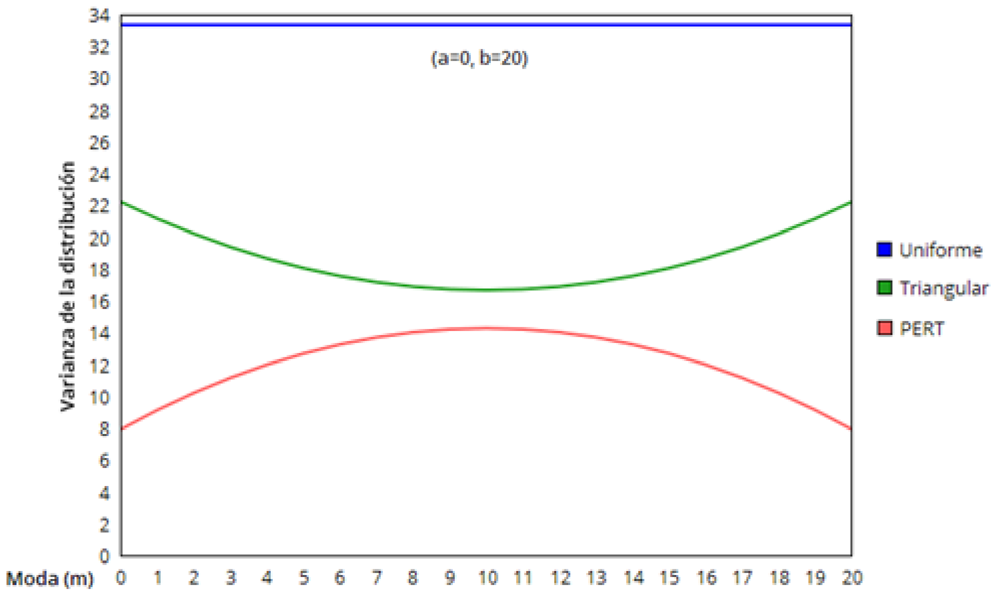

Figura 3.4: Varianzas de las distribuciones uniforme, triangular y PERT como funciones de la moda ($a = 0, b = 20$). Fuente: Elaboración propia

En la figura 3.4 se puede observar que la varianza de la distribución triangular siempre es menor a la de la distribución uniforme, y la varianza de la distribución PERT siempre es menor a la varianza de las otras dos distribuciones, sin importar el valor de la moda. También puede apreciarse que cuando $m = 10$, la distribución triangular alcanza su mínimo, mientras que la distribución PERT alcanza su máximo.

En general, cuando la moda de una variable aleatoria con distribución triangular o PERT es igual al punto medio entre el mínimo y el máximo, es decir $(a + b)/2$, la esperanza de la variable coincide con la moda de la misma. Para convencerse de lo anterior, obsérvese la siguiente prueba.

Sean $X \sim Triangular\left(a, b, \frac{a+b}{2}\right)$ y $Y \sim PERT\left(a, b, \frac{a+b}{2}\right)$

$$\Rightarrow E[X] = \frac{a + b + \frac{a+b}{2}}{3} = \frac{2a + 2b + a + b}{2 \times 3} = \frac{3a + 3b}{2 \times 3} = \frac{a + b}{2}$$

$$\Rightarrow E[Y] = \frac{a + b + 4\left(\frac{a+b}{2}\right)}{6} = \frac{a + b + 2a + 2b}{6} = \frac{3a + 3b}{6} = \frac{a + b}{2}$$

Además, cuando la media y la moda coinciden, las funciones de densidad de las distribuciones PERT y triangular son simétricas, y como se observa en la ilustración de la parte inferior, en la distribución PERT, los valores más cercanos a los extremos tienen menor probabilidad de ocurrencia que en la distribución triangular.

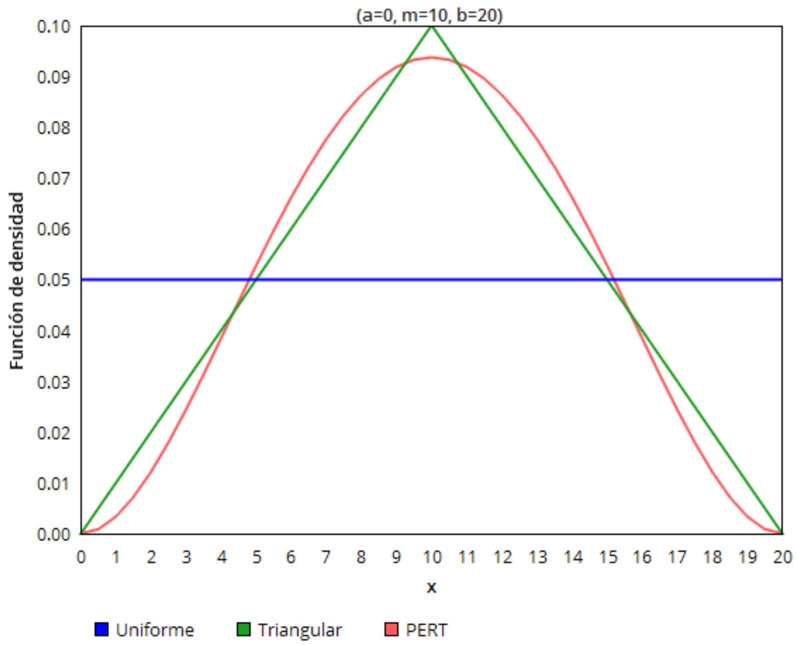

Figura 3.5: Funciones de densidad de las distribuciones uniforme, triangular y PERT $(a = 0, b = 20, m = 10)$. Fuente: Elaboración propia

Capítulo 4

Modelos de Regresión

Una técnica estadística útil para el desarrollo de ecuaciones matemáticas que muestran la relación entre variables es el análisis de regresión. En términos de esta metodología, se conoce como variable de respuesta, variable a explicar, variable endógena o variable dependiente a la variable que se desea predecir. Aquellas variables que se utilizan para predecir los valores de la variable de respuesta se les conoce como variables explicativas, variables endógenas, variables independientes, o variables de entrada.

El modelo más simple de regresión relaciona dos variables: una endógena y una exógena de forma lineal,

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad (4.0.1)$$

donde:

- Y es la variable endógena, es decir, la variable que se desea explicar.
- X es la variable exógena, es decir, la variable explicativa.
- El intercepto β_0 y la pendiente β_1 son los coeficientes del modelo de regresión. Si se define K como el número de pendientes que se estimarán en el modelo, se tienen $K = 1$ pendientes a estimar.
- e es la variable aleatoria de error, o perturbación.
- El subíndice i representa el número de observación.
- n es el tamaño de la muestra, o bien, el número de observaciones de las variables (Y, X) que se tienen disponibles.

Existen varias razones por las cuales se introduce la variable de error e_i, por ejemplo:

- Errores de medida producidos al recopilar los datos.
- Errores ocasionados al omitir factores que pueden fungir como variables exógenas.
- Factores no predecibles provocados por el contexto de las variables, o efectos no cuantificables.

Lo anterior hace referencia a modelos en que se cuenta con una sola variable explicativa. Sin embargo, en muchas situaciones, la variable de respuesta es dependiente de una multitud de variables exógenas.

Cuando se cuenta con un solo una variable explicativa y la relación entre ésta y la variable de respuesta se aproxima a una línea recta, al análisis involucrado se le conoce como regresión lineal simple. Si están involucradas dos o más variables explicativas en un modelo de regresión, se trata de un modelo de regresión múltiple.

El modelo de regresión lineal múltiple supone que la variable de respuesta Y está descrita por las variables explicativas $X_j, \ \ j = 1, \dots, k$, a través de la relación

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + e \qquad (4.0.2)$$

En esta expresión, $\beta_0, \beta_1, \dots, \beta_k$ son los parámetros de regresión y e es una variable aleatoria de error cuya media es 0. Los parámetros de regresión no serán conocidos inicialmente y deben ser estimados de un conjunto de datos.

Supóngase que se dispone de un conjunto de n respuestas correspondientes a n diferentes conjuntos de los k valores de entrada. Si y_i denota la i-ésima respuesta, y los k valores de entrada correspondientes a esta respuesta sean $x_{i1}, \dots, x_{ik}, i = 1, \dots, n$. Así, por ejemplo, y_i fue la respuesta cuando los k valores de entrada fueron $x_{i1}, \dots, x_{ik}$. Utilizando (4.0.2), se puede escribir lo anterior en forma matricial,

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} \qquad (4.0.3)$$

En adelante, la variable x_0 tendrá el valor 1 para cada observación. Escrito de otra forma, $x_{i0} = 1, \forall i \in \{1, \dots, n\}$. Denótense las matrices anteriores como sigue:

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \qquad e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

De esa forma, sustituyendo (4.0.4) en (4.0.3),

$$Y = X\,\beta + e \qquad (4.0.4)$$

Una regresión consiste en la estimación de los parámetros $\beta_j, j = 0,1, \dots, k$ a partir de una muestra. La función de regresión muestral (FRM) está determinada por medio de esa muestra, y está denotada de la siguiente manera:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik} \qquad (4.0.5)$$

Esta función, permite calcular el valor ajustado $\hat{y}_i$ del valor real y_i, y en ella, $\hat{\beta}_0$, $\hat{\beta}_1$, …, $\hat{\beta}_k$ son los estimadores de los parámetros $\beta_0, \beta_1, \dots, \beta_k$.

A la diferencia entre el valor muestral de una observación y_i y su correspondiente valor ajustado $\hat{y}_i$, se le conoce como residuo $\hat{e}_i$.

$$\hat{e}_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}) \qquad (4.0.6)$$

Esta función también puede ser escrita en forma matricial,

$$\hat{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}, \qquad \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}, \qquad \hat{e} = \begin{bmatrix} \hat{e}_1 \\ \vdots \\ \hat{e}_n \end{bmatrix} \qquad (4.0.7)$$

Por lo cual, el modelo ajustado que engloba todas las observaciones de la muestra es el siguiente:

$$\hat{Y} = X\hat{\beta} \qquad (4.0.8)$$

Mientras que, el vector de los residuos es igual a la diferencia entre el vector de los valores poblacionales y el vector de los valores ajustados.

$$\hat{e} = Y - \hat{Y} = Y - X\hat{\beta} \qquad (4.0.9)$$

4.1 Estimación por el método de mínimos cuadrados

En general, es deseable que los residuos, es decir, la diferencia entre las observaciones de la variable de respuesta y su correspondiente valor ajustado en la recta ajustada, tengan valores pequeños, pero, resulta complicado lograr que cada esto suceda si se trata cada residuo de forma individual, una forma práctica de minimizar estas distancias es minimizando la suma de las mismas. Sin embargo, con el fin de evitar que se anulen entre ellos, es necesario utilizar el valor absoluto de los residuos, no obstante, el manejo de valores absolutos en las técnicas de optimización, es tedioso y poco práctico. Una alternativa a la suma de los valores absolutos de los residuos, es la suma de los cuadrados de los residuos, ya que ambas funciones alcanzan su mínimo en los mismos valores. Al procedimiento de estimación descrito se le conoce como método de mínimos cuadrados.

Defínase lo siguiente:

$$SEC = \sum_{i=1}^{n} \hat{e}_i^2 = \sum_{i=1}^{n} \left(yi - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}\right)^2 \qquad (4.1.1)$$

Para minimizar SEC es necesario obtener sus derivadas parciales con respecto a las variables $\hat{\beta}_0$, $\hat{\beta}_1$, …, $\hat{\beta}_k$, tal como se muestra a continuación:

$$\frac{\partial SEC}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n}\left(yi - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}\right) = 0$$

$$\frac{\partial SEC}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n}\left(yi - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}\right)x_{i1} = 0$$

$$\vdots$$

$$\vdots$$

$$\frac{\partial SEC}{\partial \hat{\beta}_k} = -2\sum_{i=1}^{n}\left(yi - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}\right)x_{ik} = 0$$

si se divide por la constante -2 en ambos lados de cada ecuación, se obtiene lo siguiente:

$$\frac{\partial SEC}{\partial \hat{\beta}_j} = \sum_{i=1}^{n}\left(yi - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}\right)x_{ij} = 0, \;\; j \in \{0, 1, \ldots, k\} \qquad (4.1.2)$$

$$\Rightarrow \sum_{i=1}^{n} yi\, x_{ij} - \sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}\right) x_{ij} = 0, \;\; j \in \{0, 1, \ldots, k\}$$

$$\Rightarrow \sum_{i=1}^{n} yi\, x_{ij} - \sum_{r=0}^{k}\sum_{i=1}^{n} \hat{\beta}_r x_{ir} x_{ij} = 0, \;\; j \in \{0, 1, \ldots, k\}$$

$$\Rightarrow \sum_{i=1}^{n} yi\, x_{ij} = \sum_{r=0}^{k}\sum_{i=1}^{n} \hat{\beta}_r x_{ir} x_{ij}\,, \;\; j \in \{0, 1, \ldots, k\} \qquad (4.1.3)$$

Utilizando la notación de (4.0.7), es posible escribir las ecuaciones que deriven de (4.1.3) en forma matricial,

$$X^T Y = X^T X \hat{\beta} \qquad (4.1.4)$$

Dado que ambos miembros de la ecuación son matrices con 1 columna y $k + 1$ filas, y la matriz $(X^T X)^{-1}$ cuenta con $k + 1$ columnas y $k + 1$ filas, es posible realizar producto matricial por la izquierda en ambos lados de la ecuación con $(X^T X)^{-1}$. Por lo que se puede obtener lo siguiente:

$$(X^T X)^{-1}(X^T Y) = (X^T X)^{-1}\left(X^T X \hat{\beta}\right)$$

$$\Rightarrow (X^T X)^{-1}(X^T Y) = (X^T X)^{-1}(X^T X)\hat{\beta}$$

$$\Rightarrow (X^T X)^{-1}(X^T Y) = \hat{\beta}$$

$$\therefore \hat{\beta} = (X^T X)^{-1}(X^T Y) \tag{4.1.5}$$

Finalmente, ya que SCE puede ser vista como una función cuadrática de los estimadores $\hat{\beta}_j$, cuyo coeficiente del término cuadrático es positivo, se puede afirmar que dicha función alcanza su mínimo en el vector $\hat{\beta} = (X^T X)^{-1}(X^T Y)$. Por lo tanto, se dice que los estimadores del vector $(X^T X)^{-1}(X^T Y)$ son los estimadores de mínimos cuadrados del modelo de regresión lineal múltiple.

4.2 Modelo de Regresión Polinomial

En ocasiones, la variable dependiente Y puede depender de un solo factor cuantitativo X, sin embargo, el modelo lineal puede representar un ajuste muy pobre. Para solucionar esto, es posible construir otra clase de modelos estadísticos, en los que ambas variables estén relacionadas de una forma no lineal, uno de estos modelos alternativos consiste en ajustar los datos mediante polinomios, donde se involucren potencias de la variable independiente, como se muestra a continuación:

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x^j + e = \sum_{j=0}^{k} \beta_j x^j + e \tag{4.2.1}$$

donde k representa el grado del polinomio y β_j el coeficiente asociado a la j-ésima potencia de la variable X.

Si $x_j = x^j,\ j \in \{1, \dots, k\}$, entonces el modelo es un modelo de regresión lineal múltiple con k variables explicativas $x_1, \dots, x_k$. Así, el modelo (4.0.4) $Y = X\beta + e$ incluye al modelo de regresión polinomial. De esa forma, las técnicas para ajustar un modelo de regresión lineal múltiple pueden ser usadas para ajustar un modelo de regresión polinomial, y, por lo tanto, los estimadores de la ecuación (4.1.5) son los mismos para esta transformación. Por ejemplo:

Supóngase que se tiene una muestra de 40 observaciones para las variables X y Y, y se ha determinado que Y depende de X de acuerdo con un polinomio de segundo orden, es decir, $y = \beta_0 + \beta_1 x + \beta_2 x^2 + e$. Además, se cuenta con la siguiente información:

$$\sum_{i=1}^{40} x_i = 200, \qquad \sum_{i=1}^{40} x_i^2 = 1020.094, \qquad \sum_{i=1}^{40} x_i^3 = 5302.578, \qquad \sum_{i=1}^{40} x_i^4 = 28056.107,$$

$$\sum_{i=1}^{40} y_i = 1025.106, \qquad \sum_{i=1}^{40} x_i y_i = 5298.42, \qquad \sum_{i=1}^{40} x_i^2 y_i = 27888.63$$

Dado lo anterior, es posible construir las matrices correspondientes para generar los estimadores de la ecuación (4.1.5),

$$X^T X = \begin{bmatrix} 40 & 200 & 1020.094 \\ 200 & 1020.094 & 5302.578 \\ 1020.094 & 5302.578 & 28056.107 \end{bmatrix}, \quad (X^T X)^{-1} = \begin{bmatrix} 74.038 & -29.775 & 2.936 \\ -29.775 & 12.03 & -1.191 \\ 2.935 & -1.191 & 0.118 \end{bmatrix},$$

$$X^T Y = \begin{bmatrix} 1025.106 \\ 5298.42 \\ 27888.635 \end{bmatrix}$$

De esa forma, el vector de estimadores está dado por la siguiente ecuación:

$$\hat{\beta} = (X^T X)^{-1}(X^T Y) = \begin{bmatrix} 74.038 & -29.775 & 2.936 \\ -29.775 & 12.03 & -1.191 \\ 2.935 & -1.191 & 0.118 \end{bmatrix} \begin{bmatrix} 1025.106 \\ 5298.42 \\ 27888.635 \end{bmatrix} = \begin{bmatrix} 3.283 \\ 0.2147 \\ 0.834 \end{bmatrix}$$

Finalmente, la ecuación de regresión está determinada de la siguiente forma:

$$y = 0.834x^2 + 0.2147x + 3.283 + e.$$

Capítulo 5

Simulación Monte Carlo

La simulación Monte Carlo es un tipo de simulación que depende de repetidas muestras aleatorias y análisis estadístico para calcular los resultados. Este método de simulación está muy estrechamente relacionado con los experimentos aleatorios, experimentos para los cuales el resultado específico no se conoce previamente.

Se hace uso de modelos matemáticos en distintas ciencias y disciplinas para describir las interacciones en un sistema utilizando expresiones matemáticas. Estos modelos comúnmente dependen de un número de parámetros de entrada, que cuando son procesados a través de las fórmulas matemáticas en el modelo, derivan en uno o más resultados.

Los parámetros de entrada para los modelos dependen de varios factores externos. A causa de estos factores, los modelos realísticos están sujetos a riesgos de la variación sistemática de los parámetros de entrada. Un modelo determinístico, el cual no considera estas variaciones, es también llamado un caso base, ya que los valores de estos parámetros de entrada son sus valores más probables. Un modelo efectivo debe tomar en consideración los riesgos asociados con diversos parámetros de entrada. In más circunstancias, los experimentadores desarrollan muchas versiones de un modelo, en las cuales pueden incluir el caso base, los mejores escenarios posibles, y el peor escenario posible para los valores de las variables de entrada.

El anterior enfoque tiene varias desventajas. Primero, puede ser difícil evaluar el mejor y peor escenario para cada variable de entrada. Segundo, todas las variables de entrada podrían no estar en su mejor o peor nivel simultáneamente. La toma de decisiones tiene a ser muy difícil, una vez que se está considerando más de un escenario. Además, a medida que un experimentador incrementa el número de casos a considerar, el versionado y almacenamiento de modelos se vuelve difícil.

En la simulación Monte Carlo, se identifica una distribución estadística la cuan puede usarse como la fuente de cada uno de los parámetros de entrada. Después, se realizan muestras aleatorias de cada distribución, que entonces representan los valores de entrada de las variables. Para cada conjunto de parámetros de entrada, se tiene un conjunto de parámetros de salida. El valor de cada parámetro de salida es un particular escenario en una la corrida de la simulación. Se colectan esos valores de salida de un número de simulaciones. Finalmente, se realiza el análisis estadístico en los valores de los parámetros de salida, para tomar decisiones. Pueden usarse las estadísticas muestrales de los parámetros de salida para caracterizar la variación de salida.

5.1 Terminología

Definición 5.1. *Muestra aleatoria*. En estadística, un subconjunto finito de individuos de una población es llamado una muestra. En una muestra aleatoria, como el nombre lo indica, las muestras son tomadas aleatoriamente de la población, lo cual implica que cada unidad de la población tiene una probabilidad igual de ser incluida en la muestra.

Definición 5.2. *Generador de números aleatorios (GNA)*. Un generador de números aleatorios es un dispositivo computacional o físico diseñado para generar una secuencia de números que parecen ser muestras independientes de una población, y que además cumplen con una serie de pruebas estadísticas. Éstos también son llamados Generadores de números pseudo-aleatorios, ya que los números aleatorios generados a través de este método no son reales, sino simulados.

5.2 Generación de Variables Aleatorias

Después de identificar las distribuciones subyacentes para los parámetros de entrada de un modelo de simulación, se generan números aleatorios a partir de estas distribuciones. Los números aleatorios generados representan valores específicos de la variable. Por ejemplo, se ha determinado que la distribución normal es la mejor para ajustar los pesos de los estudiantes de una escuela. Si se quiere usar esta información en un modelo que tiene los pesos como un parámetro de entrada, debería generarse un número aleatorio de esta distribución, y usar ese número como un peso representativo.

En esta sección, se describirán dos métodos para la generación de variables aleatorias discretas y continuas. De igual forma, se ejemplificará la forma en que se generan valores aleatorios para algunas de las distribuciones descritas en el capítulo 3. Finalmente, en los anexos se exponen los códigos de VBA para generar valores aleatorios en la hoja de cálculo de acuerdo con los métodos de las secciones 5.2.1 – 5.2.6.

5.2.1 Método de la Transformación Inversa

El método de la transformación inversa proporciona la ruta más directa para la generación de una muestra aleatoria de una distribución. En este método, se usa la inversa de la función de distribución acumulada, y convierte un número aleatorio entre 0 y 1 a un valor aleatorio de la distribución de entrada. El proceso puede ser descrito matemáticamente como sigue.

Sea X una variable aleatoria continua, de la cual se desea generar valores, y sean f su función de densidad de probabilidad, y F su función de distribución acumulada, la cual es continua y estrictamente creciente en el intervalo $(0,1)$. Denótese con F^{-1} a la función inversa de F, la cual es también llamada función de distribución acumulada inversa. Entonces, los siguientes dos pasos generarán un número aleatorio X de la *fdp* f.

$$\begin{aligned} &1. \, Generar \; U \sim U(0,1). \\ &2. \, Calcular \; X = F^{-1}(U). \end{aligned} \tag{5.2.1}$$

Nótese que, como $0 \leq U \leq 1$, $F^{-1}(U)$ existe siempre.

El método de la transformación inversa también puede ser usado cuando X es una variable aleatoria discreta. Para distribuciones discretas, si p_i es la función de masa de probabilidad, la función de probabilidad acumulada está dada por:

$$F(x) = P\{X \leq x\} = p_x + \sum_{i \leq x} p_i \tag{5.2.2}$$

La función de probabilidad acumulada es una función con saltos discretos. Entonces, el segundo paso del algoritmo mencionado anteriormente para la generación de variables aleatorias de una distribución continua puede ser reemplazado por el siguiente: determinar el entero positivo más pequeño I tal que $U \leq F(x_I)$, y devuelva $X = x_I$.

Una importante ventaja del método de la transformación inversa es que puede ser usado para generar números aleatorios de una distribución truncada. Este método puede ser usado para cualquier tipo de función de distribución, incluyendo funciones que son una mezcla de distribuciones discretas y continuas. Una desventaja proviene del hecho de que este método se vuelve difícil de implementar si no hay una forma cercana a la función de distribución acumulada para una distribución.

5.2.2 Método de Aceptación y Rechazo

Muy comúnmente, una forma analítica para $F(x)$ es desconocida o demasiado compleja para trabajar con ella, así que el obtener una función inversa de dicha función no es práctico. Supóngase que para algún valor dado de x, la función de densidad de probabilidad $f(x)$ puede ser calculada, y, además se tiene suficiente información de $f(x)$ para encerrarla completamente dentro de una figura que es C veces una distribución fácilmente generada.

Frecuentemente $h(x)$ es uniforme o es una suma normalizada de distribuciones uniformes. Nótese que tanto $f(x)$ como $h(x)$ deben ser normalizadas a la unidad de área. Para generar $f(x)$, primero se genera un candidato x de acuerdo con $h(x)$. Calcular $f(x)$ y la altura de la figura $Ch(x)$; generar U y probar si $UCh(x) \leq f(x)$. Si se cumple la condición anterior, el punto x es aceptado y servirá como elemento para la muestra aleatoria; de lo contrario, el valor x será rechazado para utilizarse en la muestra, y deben generarse nuevos valores hasta que alguna cumpla la condición. Si se considera a x como la abscisa y a $UCh(x)$ como la ordenada de un punto en una gráfica bidimensional, estos puntos poblarán el área $Ch(x)$ completa; entonces se aceptarán los puntos que caigan bajo $f(x)$.

Resumiendo, el método de aceptación y rechazo se realiza mediante los siguientes pasos:

1. Generar un valor aleatorio x distribuido de acuerdo con h.
2. Generar un valor aleatorio U (independiente de x).
3. Si $UCh(x) \leq f(x)$ el valor x se acepta como parte de la muestra, de otra forma se rechaza.

El procedimiento anterior se utiliza cuando la función de densidad f de la que se desean simular valores, cuenta con un soporte no necesariamente finito. Existe un caso particular de este método para generar valores de variables aleatorias cuya función de densidad tiene un soporte acotado (a, b), algunas distribuciones que caen en este caso son la distribución triangular, la distribución beta y la distribución PERT. Este método se describe a continuación.

Sea C un valor constante tal que: $C \geq \max\{f(x): x \in (a, b)\}$. El método consiste en generar puntos aleatorios cuyas abscisas estarán distribuidas uniformemente en el intervalo (a, b) y cuyas ordenadas se encontrarán distribuidas uniformemente en el intervalo $(0, C)$, de esa forma, los puntos se encontrarán dentro del rectángulo $(a, b) \times (0, C)$, y aquellos puntos que se encuentren bajo la curva de la función de densidad serán aceptados, o de lo contrario, serán rechazados. Si un punto es aceptado, el valor de su ordenada se tomará como un valor de la variable aleatoria con función de densidad f.

5.2.3 Generación de muestras aleatorias uniformes en el intervalo (a, b)

Es bien sabido que algunos paquetes de software como las hojas de cálculo, y algunos dispositivos como las calculadoras científicas tienen la capacidad de generar valores aleatorios o pseudo-aleatorios distribuidos uniformemente en el intervalo $(0,1)$, y como puede observarse en las secciones anteriores, la generación de los valores con esta distribución es el primer paso para realizar los procedimientos presentados anteriormente. Si bien es posible generar valores aleatorios de una densidad f por más de un método, para cada tipo de distribución existe un método que resulta más sencillo que los demás al ponerlo en práctica.

Una forma sencilla de generar valores aleatorios de una distribución uniforme en el intervalo (a,b) es por medio del método de la transformación inversa, y de acuerdo con lo propuesto en la sección 5.2.1 consiste en hallar una función inversa para la función de distribución, generar un valor aleatorio uniforme en el intervalo $(0,1)$ y evaluar este valor en la función inversa encontrada. Este procedimiento se muestra desarrollado a continuación.

La función de distribución de una variable aleatoria $U(a,b)$ está representada por

$$F(x) = \frac{x-a}{b-a}, \qquad a < x < b$$

según la ecuación (3.2.2). Por lo que la función de distribución inversa $F^{-1}(x)$ está dada de la siguiente manera:

$$F^{-1}(u) = (b-a)u + a, \qquad 0 < u < 1 \qquad (5.2.3)$$

Por ejemplo, supóngase que se desea obtener una muestra aleatoria de 5 valores distribuidos uniformemente en el intervalo $(2,5)$ a partir de los valores 0.3953, 0.0542, 0.2019, 0.389, 0.3384. Sustituyendo estos valores, y los de los parámetros $a = 2, b = 5$ en la ecuación (5.2.3) se obtienen los números aleatorios que se muestran en la siguiente tabla:

Valor $U(0,1)$	Valor $U(2,5)$ o $F^{-1}(u)$
0.3953	3(0.3953)+2=3.1859
0.0542	3(0.0542)+2=2.1626
0.2019	3(0.2019)+2=2.6057
0.389	3(0.389)=3.167
0.3384	3(0.3384)=3.0152

Tabla 5.1: Generación de valores distribuidos uniformemente en el intervalo $(2,5)$.
Fuente: Elaboración propia

Así, los valores uniformemente distribuidos en el intervalo $(2,5)$ generados son: 3.1859, 2.1626, 2.6057, 3.167, 3.0152.

5.2.4 Generación de muestras aleatorias con distribución Bernoulli

Debido a la poca complejidad que muestran la función de densidad y la interpretación de la distribución Bernoulli, un método sencillo para la generación de valores con esta distribución es el de la transformación inversa. Además, esa sencillez es heredada por su función de distribución, la cual, de acuerdo con la ecuación (5.2.2) está denotada de la siguiente manera:

$$F(x) = \begin{cases} 1-p, & si\ x = 0 \\ 1, & si\ x = 1 \end{cases} \tag{5.2.4}$$

Aunque hallar la función de distribución inversa resultara ser lo más lógico, es importante notar que en una variable aleatoria U con distribución uniforme en el intervalo $(0,1)$, la probabilidad de que $0 < U \leq 1-p$ es igual a $1-p$, y la probabilidad de que $1-p < U < 1$ es igual a p, es decir, $P\{0 < U \leq 1-p\} = 1-p$, y $P\{1-p < U < 1\} = p$. Haciendo uso de lo anterior, es posible construir la siguiente regla:

$$X = \begin{cases} 0, & si\ 0 < u \leq 1-p \\ 1, & si\ 1-p < u < 1 \end{cases} \tag{5.2.5}$$

donde u pertenece a la distribución $U \sim U(0,1)$. Con las condiciones de (5.2.5) se puede asegurar que $P\{X=0\} = 1-p$, y que $P\{X=1\} = p$.

A manera de ejemplo, supóngase que se desea simular una muestra aleatoria de una distribución Bernoulli con parámetro $p = 0.7$. Para generar los valores dicotómicos, primero deben generarse los valores uniformemente distribuidos entre 0 y 1, cuando un valor sea menor o igual que 1-0.7=0.3 se agregará un 0 a la muestra, mientras que, cuando un valor uniforme exceda a 0.3 se agregará un 1 a la muestra aleatoria.

5.2.5 Generación de muestras aleatorias con distribución Triangular

En esta sección se presentará la forma de generar valores aleatorios con distribución triangular por medio de los dos métodos presentados en las secciones anteriores; el método de la transformación inversa y el método de aceptación y rechazo.

Comenzando con el método de la transformación inversa. Supóngase que se desea generar una muestra aleatoria de la variable X con distribución triangular, y parámetros $a, b\ y\ m$. De acuerdo con la ecuación (3.3.2) la función de distribución acumulada de X está dada por:

$$F_x(x) = \begin{cases} \dfrac{(x-a)^2}{(b-a)(m-a)}, si\ a < x \leq m \\ 1 - \dfrac{(b-x)^2}{(b-a)(b-m)}, si\ m < x < b \end{cases}$$

Es claro que $F_X(m) = (m - a)/(b - a)$, por lo que la función de distribución inversa correspondiente a la distribución triangular es la siguiente:

$$F_X^{-1}(u) = \begin{cases} \sqrt{u(b-a)(m-a)} + a, & si\ 0 < u \leq \dfrac{m-a}{b-a} \\ b - \sqrt{(1-u)(b-a)(b-m)}, & si\ \dfrac{m-a}{b-a} < u < 1 \end{cases} \tag{5.2.6}$$

Una vez generado un valor u de acuerdo con la distribución uniforme con parámetros 0 y 1, puede transformarse en un valor $x = F_X^{-1}(u)$ distribuido triangularmente.

Para continuar con el método de aceptación y rechazo es necesario hacer referencia a la ecuación (3.3.1), que representa la función de densidad de una variable aleatoria triangular con parámetros $a, b\ y\ m$, y está denotada como sigue:

$$f_X(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(m-a)}, & si\ a \leq x < m \\ \dfrac{2}{b-a}, & si\ x = m \\ \dfrac{2(b-x)}{(b-a)(b-m)}, & si\ m < x \leq b \end{cases}$$

Así que, para generar un valor aleatorio triangular por el método de aceptación y rechazo, primero debe generarse un valor aleatorio u_1 distribuido uniformemente en el intervalo (a, b), y otro valor u_2 distribuido uniformemente en el intervalo $(0, 2/(b - a))$, ya que $\max\{f_X(x): a \leq x \leq b\} = 2/(b - a)$. Estos valores pueden ser generados por medio del método de transformación inversa. La siguiente regla determina si un valor sirve como elemento para la muestra deseada:

$$\begin{aligned} &- Si\ u_2 \leq f_X(u_1), el\ valor\ u_1\ se\ acepta\ como\ parte\ de\ la\ muestra. \\ &- Si\ u_2 > f_X(u_1), el\ valor\ u_1\ debe\ ser\ rechazado. \end{aligned} \tag{5.2.7}$$

5.2.6 Generación de muestras aleatorias con distribución PERT

Debido a la complejidad de la función de distribución acumulada de la distribución PERT, resulta poco práctica la simulación de valores aleatorios con esta distribución haciendo uso del método de transformación inversa. Por consiguiente, en este trabajo se explicará la simulación de valores aleatorios con distribución PERT por medio del método de aceptación y rechazo.

Sea X una variable aleatoria con distribución PERT y parámetros a, b y m, de la cual se desea generar una muestra aleatoria. A diferencia de la distribución triangular, el valor máximo que puede tomar la función de densidad depende de la moda m, y no sólo de los valores de a y b, además, la forma en que está representado dicho valor es mucho más compleja, razón por la cual, en adelante será denotado por $f_X(m)$, siendo $f_X(m)$ según la ecuación (3.5.5). Para fines prácticos, el valor de $f_X(m)$ puede ser determinado con la ayuda de un método numérico o algún paquete de software.

De la misma forma que se hace con la distribución triangular, en el caso de la distribución PERT, deben generarse dos valores aleatorios u_1 de acuerdo con la distribución uniforme en el intervalo (a, b), y u_2 de acuerdo con la distribución uniforme en el intervalo $(0, f_X(m))$. Para la aceptación o el rechazo de un valor, puede utilizarse la misma regla que en el caso triangular, es decir (5.2.7).

En el siguiente ejemplo, se muestra una serie de diez valores candidatos a formar parte de una muestra aleatoria con distribución $PERT(10,20,15)$. Para este caso $f_X(m) = 0.1875$.

$u_1 \sim U(10, 20)$	$u_2 \sim U(0, 0.1875)$	$f_X(u_1)$	*Decisión*
19.5382	0.0603	0.0058	Rechazado
19.0090	0.1205	0.0239	Rechazado
17.0228	0.0765	0.1311	Aceptado
11.9554	0.1874	0.0742	Rechazado
11.0384	0.1319	0.0260	Rechazado
14.0900	0.1538	0.1753	Aceptado
16.7699	0.0418	0.1435	Aceptado
19.1251	0.1481	0.0191	Rechazado
12.8473	0.0201	0.1244	Aceptado
17.5611	0.1129	0.1020	Rechazado

Tabla 5.2: Generación de valores aleatorios con distribución PERT(10,20,15) a través del método de aceptación y rechazo. Fuente: Elaboración propia

Como puede observarse en la tabla 5.2, de los diez valores generados, sólo cinco de ellos fueron aceptados para integrar la muestra aleatoria, esto indica que, en ocasiones es necesario generar una cantidad de números aleatorios mayor al doble del tamaño deseado de la muestra.

Capítulo 6

Caso de Estudio

El Centro Avanzado de Propulsión ubicado en la ciudad de Coventry (al cual, en adelante se le denotará por CAP), tiene el papel de posicionar al Reino Unido como un centro de excelencia para el desarrollo y producción de proyectos de propulsión con baja emisión de carbono.

El CAP es una empresa privada de responsabilidad limitada, una colaboración amplia entre los productores y los innovadores de sistemas de propulsión con baja emisión de carbono. Una de las tareas de este centro es facilitar las asociaciones entre los que tienen buenas ideas y los que pueden llevarlas al mercado. El CAP impulsa el desarrollo de proyectos que otorguen a las partes involucradas un crecimiento sostenible y rentable.

La cobertura de los proyectos del CAP incluye cuatro de las cinco tecnologías estratégicas identificadas por el Consejo Automotriz de ese país y presentadas en el Evento de Vehículos de baja emisión de carbono en Millbrook en septiembre de 2013:

- Máquinas eléctricas y electrónicos de potencia
- Almacenamiento y gestión energética
- Sistemas de propulsión térmica
- Vehículos de peso ligero y estructuras de trenes motores

La quinta estrategia tecnológica; Movilidad Inteligente compete a la Catapulta de Sistemas de Transporte y al Centro para Vehículos Conectados y Autónomos.

Cada uno de los proyectos respaldados por el CAP integra al menos una de las cuatro tecnologías a través de su actividad de investigación, el desarrollo y la producción de tecnología de baja emisión de carbono. Los sistemas de propulsión pueden representar hasta el 50% del valor total de un vehículo, así que las actividades respaldadas por el CAP harán una significativa contribución a la prosperidad económica del Reino Unido, construyendo una nueva capacidad en la base de la oferta del Reino Unido.

Como se ha mencionado a lo largo de esta sección, uno de los mayores objetivos del CAP es desarrollar proyectos con una significativa reducción en la emisión de CO_2, en comparación con los sistemas populares, sin embargo, el mayor reto consiste en no afectar sustancialmente la efectividad del sistema de producción en serie. Para ello, es necesario contar con al menos un socio dispuesto a absorber una fracción del financiamiento de un proyecto específico, mientras que el gobierno inglés será el encargado de financiar el resto del proyecto.

El financiamiento de los proyectos no es el único aspecto que debe considerarse para lograr los propósitos establecidos. Es importante atender cautelosamente las posibles desviaciones que pueden surgir al momento de asignar presupuestos para Investigación y Desarrollo (I+D). Las desviaciones en el presupuesto ocurren de manera mensual, y el conjunto de éstas y las demás a las que es posible enfrentarse se expone a continuación:

1. *Probabilidad de ejecución del proyecto*.

   Pese a haberse planificado un proyecto no es posible que éste se lleve a cabo, por lo que existe una probabilidad de que esto ocurra, y, por consiguiente, una probabilidad de lo contrario. Estas probabilidades sólo se contemplarán antes de dar inicio al proyecto, ya que una vez que se comienza su ejecución, la probabilidad de que se lleve a cabo es igual al 100%, y cuando esto sucede, deben contemplarse las desviaciones descritas posteriormente.

2. *Posposición en la fecha de inicio del proyecto*.

   La variabilidad en el tiempo que implica la realización de trámites y papeleos burocrático para los permisos de investigación puede provocar una demora incierta con respecto a la fecha de inicio establecida para un proyecto. Aunque esta variación puede llegar a alterar las fechas de inicio y de terminación de los proyectos, no afecta las cantidades presupuestadas mensualmente, ni el monto total asignado.

3. *Prolongación del tiempo requerido para un proyecto*.

   Debido a que los materiales integrados en los productos desarrollados en el proyecto pueden estar sujetos a una disponibilidad limitada en un periodo de tiempo, existe la posibilidad de que se requiera un mayor número de meses al que se contempló inicialmente para la terminación de un proyecto. Aunque esto no altera el monto total a gastar del proyecto establecido inicialmente, causará una variación en los presupuestos de los meses restantes del proyecto. Una forma efectiva de asignar estos presupuestos cumpliendo la condición de no alterar el monto total se muestra en seguida.

   Cuando el mes de referencia es un mes transcurrido, evidentemente no es posible modificar el presupuesto para ese periodo. En cambio, cuando el periodo aún no ha dado inicio, la fórmula para calcular el monto gastado de dicho periodo es la siguiente:

$$P_i = PI_i\left(\frac{m_p - m_t}{m_p - m_t + m_e}\right), \qquad m_t < i \le m_p \tag{6.0.1 a}$$

donde P_i: Monto calculado para el periodo i cuando existe una extensión en el número de meses
PI_i: Presupuesto asignado inicialmente al periodo i
m_p: Número de meses proyectados inicialmente
m_t: Número de meses transcurridos desde que se dio inicio al proyecto
m_e: Variable aleatoria que representa el número de meses de extensión del proyecto.

Para los meses posteriores a los proyectados inicialmente, el presupuesto puede asignarse de la siguiente manera:

$$P_i = \frac{PT - \sum_{j=1}^{m_t} PI_j - \sum_{k=m_t+1}^{m_p} P_k}{m_e} \tag{6.0.1 b}$$

donde PT: Presupuesto total asignado al proyecto
P_k: Es calculado de acuerdo con la formula (6.0.1 a)
Y PI_j, m_t, m_p y m_e son como se describieron anteriormente.

Un ejemplo para la asignación de los presupuestos se esquematiza en la tabla 6.1.

Mes	Presupuesto Inicial	Escenario Posible
1	£454,214.72	£454,214.72
2	£278,665.60	£278,665.60
3	£278,665.60	£167,199.36
4	£199,760.00	£119,856.00
5	£199,760.00	£119,856.00
6		£135,637.12
7		£135,637.12

Tabla 6.1: Ejemplo de la asignación de presupuestos cuando existe una extensión de dos meses requeridos para un proyecto. Fuente: Elaboración propia

En la tabla anterior, el presupuesto de los dos primeros meses no se ve afectado debido a que éstos ya han transcurrido, mientras que del tercero al quinto mes se han multiplicado por 3/5; valor calculado de acuerdo con la ecuación (6.0.1 a), y los presupuestos del sexto y séptimo mes se han calculado de acuerdo con la formula (6.0.1 b).

4. *Cambios en los presupuestos mensuales de un proyecto.*

Una vez contempladas las variaciones anteriores, resulta lógico pensar que los montos presupuestales asignados mensualmente a un proyecto no serán precisamente iguales a las requeridas en la práctica. Para tales efectos, el equipo de expertos del CAP determinó que en los casos en los cuáles existen dichos cambios, una buena forma de modelarlos es por medio de una regresión polinomial de mínimos cuadrados de cuarto orden, cuya variable explicativa es el tiempo. Dicho comportamiento se define algebraicamente de la siguiente manera:

$$P(t) = \beta_4 t^4 + \beta_3 t^3 + \beta_2 t^2 + \beta_1 t + \beta_0 \qquad (6.0.2)$$

Donde $P(t)$ representa el monto gastado mensualmente en el periodo t. Una convincente razón del porqué utilizar este modelo para representar el presupuesto egresado por proyecto es el hecho de que, en los modelos de regresión por mínimos cuadrados, la suma de los valores ajustados es igual a la suma de los valores de las observaciones, lo cual, para efectos de este caso de estudio, se traduce en que el presupuesto total de un proyecto no debe verse afectado, aun cuando los presupuestos mensuales sufran las variaciones descritas.

t	Presupuesto Inicial	Escenario Posible	Escenario proyectado
	£454,214.72	£454,214.72	£454,214.72
	£278,665.60	£278,665.60	£278,665.60
1	£278,665.60	£139,332.80	139489.3587
2	£199,760.00	£99,880.00	99097.20635
3	£199,760.00	£99,880.00	101445.5873
4		£113,030.93	111465.346
5		£113,030.93	113813.727
6		£113,030.93	112874.3746

Tabla 6.2: Ejemplo de la asignación proyección de un escenario con una extensión de tres meses y variación mensual. Fuente: Elaboración propia

Supóngase que para los datos de la tabla 6.1 se presenta una extensión de 3 meses, y han transcurrido dos meses desde el inicio del proyecto. En la tabla 6.2 se presenta el escenario proyectado para los presupuestos con esta variación.

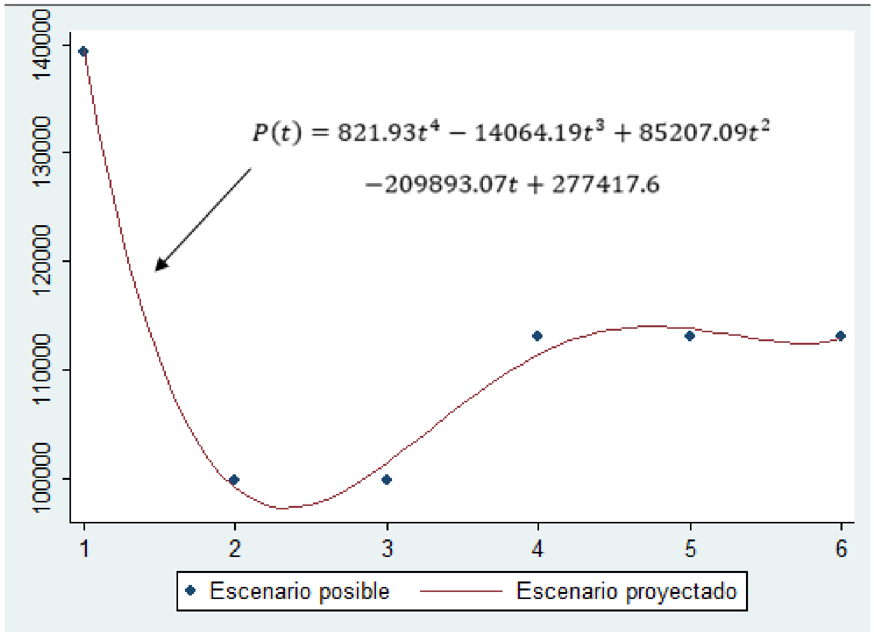

Figura 6.1: Ajuste del escenario posible con el escenario proyectado de los presupuestos de la tabla 6.2. Fuente: Elaboración propia

5. *Desviaciones con respecto a la totalidad del presupuesto.*

El último punto que debe considerarse dentro de las desviaciones radica en el hecho de que pueden presentarse situaciones en las cuales el total gastado sea superior o inferior al monto presupuestado. Al contrario del resto de las desviaciones, ésta afecta directamente al monto total por proyecto. Al construir el modelo, el porcentaje de variación debe ser ejercido de igual manera para los montos mensuales de los periodos futuros del proyecto, de forma que la variación en el total presupuestado coincida con el porcentaje arrojado por la simulación.

Utilizando la misma información que en las tablas 6.1 y 6.2, asúmase que ocurrió un retraso de un mes antes de iniciar el proyecto y existe una desviación del 5% adicional al monto total presupuestado, en la tabla 6.3 se muestra un ejemplo del prorrateo para este caso, habiendo transcurrido dos meses y una extensión de tres periodos mensuales.

t	Periodo proyectado	Presupuesto Original	Presupuesto con retraso	Presupuesto con retraso y extensión	Presupuesto con retraso, extensión y variación mensual	Presupuesto con retraso, extensión, variación mensual y desviación porcentual
1		£454,214.72				
2		£278,665.60	£454,214.72	£454,214.72	£454,214.72	£454,214.72
3		£278,665.60	£278,665.60	£278,665.60	£278,665.60	£278,665.60
4	1	£199,760.00	£278,665.60	£139,332.80	139489.3587	£154,000.78
5	2	£199,760.00	£199,760.00	£99,880.00	99097.20635	£109,406.53
6	3		£199,760.00	£99,880.00	101445.5873	£111,999.22
7	4			£113,030.93	111465.346	£123,061.36
8	5			£113,030.93	113813.727	£125,654.05
9	6			£113,030.93	112874.3746	£124,616.97

Tabla 6.3: Ejemplo del prorrateo para un proyecto con retraso en el mes de inicio, tres meses de extensión, variación mensual y variación del 5% en el total presupuestado. Fuente: Elaboración propia

6.1 Definición y construcción del modelo

La base de datos del CAP consta de 17 proyectos, los cuales están denotados por la siguiente nomenclatura: $CAPn$, (donde $1 \leq n \leq 17$), cada uno de ellos conformado a su vez por uno o más subproyectos. Supóngase que el proyecto n está conformado por M subproyectos, entonces, el subproyecto m del proyecto n está definido por $CAPn - Pm$ (donde $1 \leq m \leq M$), el número total de subproyectos es 31, y se encuentran estructurados como en la tabla 6.4.

Nombre del proyecto	Nombre del subroyecto	Nombre del proyecto	Nombre del subroyecto	Nombre del proyecto	Nombre del subroyecto	Nombre del proyecto	Nombre del subroyecto
CAP1	CAP1-P1	**CAP3**	CAP3-P2	**CAP4**	CAP4-P5	**CAP11**	CAP11-P1
	CAP1-P2		CAP3-P3		CAP4-P6	**CAP12**	CAP12-P1
	CAP1-P3		CAP3-P4	**CAP5**	CAP5-P1	**CAP13**	CAP13-P1
	CAP1-P4		CAP3-P5	**CAP6**	CAP6-P1	**CAP14**	CAP14-P1
	CAP1-P5	**CAP4**	CAP4-P1	**CAP7**	CAP7-P1	**CAP15**	CAP15-P1
CAP2	CAP2-P1		CAP4-P2	**CAP8**	CAP8-P1	**CAP16**	CAP16-P1
	CAP2-P2		CAP4-P3	**CAP9**	CAP9-P1	**CAP17**	CAP17-P1
CAP3	CAP3-P1		CAP4-P4	**CAP10**	CAP10-P1		

Tabla 6.4: Estructura y nomenclatura de los proyectos y subproyectos del CAP. Fuente: Elaboración propia

Con el fin de facilitar el análisis de la calendarización, el periodo que comprende del 1 de abril del año 2014 al 30 de abril del mismo año, en adelante, será referido como el periodo 1, al periodo del 1 de mayo de 2014 al 31 de mayo de 2014 como el periodo 2, y los siguientes periodos de forma sucesiva, siendo el periodo 157, el cual está comprendido entre el 1 de abril del año 2027 y el 30 de abril del mismo año el último periodo considerado, esto, debido a que ninguno de los proyectos analizados puede concluirse en un periodo posterior a éste.

Es de suma importancia mencionar que la fecha en que se recuperó la base de datos, y en la cual se perdió continuidad de la misma fue el día 7 de junio del año 2016, razón por la cual el periodo número 27, que corresponde al mes de junio del año 2016, será el periodo de referencia para realizar el análisis, principalmente en aquellos periodos en que ocurren las variaciones tipo 3, 4 y 5, es decir, cuando se presenta por lo menos uno de los siguientes casos:

- Existe una prolongación en el número de meses necesarios para la realización de un proyecto y deban abarcarse periodos posteriores al mes de mayo del 2016
- Para los periodos en los que pueden existir variaciones en los presupuestos mensuales.
- Cuando es necesario asignar variaciones porcentuales a los presupuestos mensuales de un proyecto.

Comprendido lo anterior, se puede proceder a explicar los conceptos englobados en la construcción del modelo. Para ello, deben definirse los dos tipos de información contenidos en el modelo.

- *Información de entrada*. Son los datos correspondientes a la base de datos proporcionada por el CAP, y en ella se incluye:

  - *Periodo de referencia*. Es el periodo en el cuál se perdió la continuidad de los datos y el punto de partida para los pronósticos subyacentes a las variaciones.

  - *Periodo/Fecha inicial estimado(a)*. Es el periodo mensual en el cual se espera que inicie un proyecto siempre que no exista algún retraso.

  - *Periodo/Fecha de fin estimado(a)*. Es el periodo mensual dentro del cual se estima que finalice un proyecto en caso de no existir algún retraso en la fecha de inicio o una extensión en el número de meses requeridos para dicho proyecto.

  - *Meses de retraso en la fecha de inicio*. Dado que se trata de un modelo de simulación, estos datos pueden referirse tanto a datos de entrada como a datos de salida, siendo los parámetros aportados los valores de entrada y siendo los valores simulados la información de salida.

Los expertos del CAP han determinado que la mejor forma de modelar esta variación es por medio de la distribución PERT con parámetros a, b y m, donde a representa el valor optimista o el tiempo mínimo de retraso, b el valor pesimista o tiempo máximo de retraso y m la moda o valor más probable. La razón de utilizar dicha distribución radica en que los valores extremos tienen una baja probabilidad de ocurrencia en comparación con los de la distribución triangular, esta afirmación se ilustra en la siguiente gráfica.

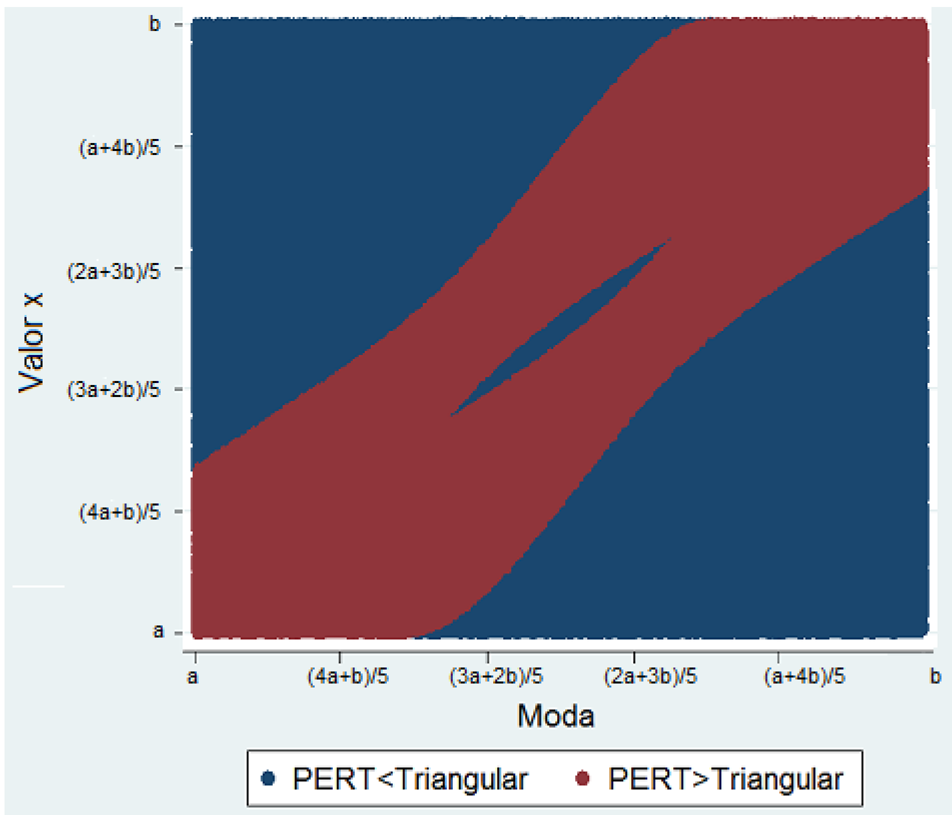

Figura 6.2: Comparación de las densidades de las distribuciones PERT y triangular de acuerdo con el valor de la moda. Fuente: Elaboración propia

La figura 6.2 consiste en un diagrama de dispersión en dónde los valores de la abscisa son las posibles modas de las distribuciones PERT y triangular con parámetros a y b, y el eje de la ordenada corresponde a los valores que puede tomar una variable aleatoria, en donde el área de color rojo representa al conjunto en el cual la densidad de la distribución PERT es mayor que la densidad de la distribución triangular y el área de color azul es el conjunto de valores de x y de la moda en donde la distribución PERT tiene una mayor densidad que la distribución triangular. En este gráfico, es fácil observar que cuando la moda es cercana al valor mínimo, los valores extremos, que en este caso son los valores cercanos al máximo, tienen una mayor densidad si se modelan con una distribución PERT que en caso de modelarse con la distribución triangular; cuando la moda es cercana al punto medio entre el mínimo y el máximo del rango, los valores más cercanos al máximo y al mínimo, así como aquellos que son bastante cercanos a la moda, tienen una menor probabilidad de ocurrencia para la distribución PERT que para la distribución triangular; por último, si la moda es cercana al máximo del rango, los valores más distantes del máximo se encuentran menos concentrados en la distribución PERT en comparación con la distribución triangular. En conclusión, es más conveniente utilizar una distribución PERT cuando se sabe que los valores extremos se presentan de manera esporádica.

- *Número de meses de extensión requeridos para la terminación del proyecto*. Al igual que el número de meses de retraso en la fecha de inicio, el número de meses de extensión involucra datos de entrada y salida, siendo los primeros la distribución asociada y sus parámetros correspondientes, de acuerdo con la experiencia de los ingenieros del CAP, los valores extremos de esta variación ocurren de manera poco frecuente, y como se mencionó anteriormente, es una propiedad de la distribución PERT, razón suficiente para hacer uso de dicha distribución (con parámetros a, b y m) al momento de modelar la extensión. Es importante señalar que tanto para el número de meses de retraso y los meses de extensión deben manejarse números enteros, por lo que los valores obtenidos por medio de la simulación deben redondearse al entero más cercano.

- *Probabilidad de iniciar el proyecto*. Anteriormente se mencionó que cada proyecto tiene una probabilidad distinta de llevarse a cabo, y una vez iniciado un proyecto, la probabilidad de ejecutarse y concluirse en un tiempo determinado es igual al 100%, por lo que sólo existen dos posibilidades: el proyecto se ejecuta o se cancela, así que la distribución ideal para modelar es la distribución Bernoulli con parámetro p, donde p es la probabilidad de ejecución del proyecto.

- *Porcentaje del presupuesto inicial ejercido*. En este caso, se cuenta únicamente con la información del porcentaje mínimo ejercido, el porcentaje máximo ejercido y el porcentaje ejercido con mayor probabilidad de ocurrencia, y de la misma forma que en los retrasos y extensiones, los valores extremos ocurren con baja frecuencia, por lo tanto, se utilizará la distribución PERT para modelar esta variación, sin embargo, dado que se trata de porcentajes, a diferencia de las variaciones expuestas anteriormente, los valores de la distribución asociada no necesariamente son enteros. La información de entrada para este caso son la distribución y los parámetros asociados, mientras que la información de salida serán los valores simulados.

- *Presupuesto original*. En este apartado se incluyen los montos presupuestados de manera mensual antes de sufrir cualquiera de las variaciones establecidas. Cada proyecto tiene un número distinto de meses, y, por consiguiente, un número distinto de montos presupuestados.

- *Información de salida*. Es el conjunto de valores obtenidos por medio de la ejecución de la simulación, y el conjunto de datos resultantes de la interacción de los mismos con aquellos proporcionados como información de entrada. La información de salida de este modelo está integrada por los siguientes puntos.

  - *Distribuciones de probabilidad simuladas*. Se trata de los valores resultantes después de efectuar la simulación, sin alteración alguna, salvo el redondeo aplicado a los meses de retraso en la fecha de inicio y a los meses de extensión requeridos para un proyecto.

  - *Periodo/Fecha de inicio*. Como se indicó anteriormente, para cada subproyecto se tiene un periodo o fecha de inicio estimado, sin embargo, en muchos casos es poco probable que el proyecto se inicie en esa fecha. Pese a no ser posible conocer el periodo de inicio preciso de un proyecto, éste puede describirse con la siguiente ecuación:

$$\begin{aligned} &\textit{Periodo de inicio} \\ &= \textit{Periodo de inicio estimado} \\ &+ \textit{Meses de retraso} \end{aligned} \tag{6.1.1}$$

    La fecha que corresponde al periodo de inicio calculado debe ser referida tal como se estableció al inicio de esta sección.

  - *Periodo/Fecha de terminación.* La fecha de inicio de un proyecto sólo se ve afectada por los retrasos, en cambio, la fecha de finalización del subproyecto puede ser retrasada según los retrasos en la fecha de inicio y los meses adicionales que requiera el proyecto. De la misma forma que con la fecha de inicio, es posible explicar la fecha de terminación en términos de las variaciones.

$$\begin{aligned} &\textit{Periodo de terminación} \\ &= \textit{Periodo de terminación estimado} \\ &+ \textit{Meses de retraso} \\ &+ \text{Meses de extensión} \end{aligned} \tag{6.1.2}$$

- *Número de periodos de duración de un proyecto*. Se trata del número de meses transcurridos desde el inicio de un proyecto hasta su terminación, y se computa de la siguiente forma:

$$\begin{aligned}&\textit{Periodos de duración}\\&= \textit{Periodo de terminación}\\&\quad - \textit{Periodo de inicio} + 1\end{aligned} \qquad (6.1.3)$$

- *Presupuesto con retraso y extensión*. Para cada uno de los periodos en los cuales se planea llevar a cabo la ejecución de un proyecto, con el fin de simplificar la estructura del modelo, es práctico realizar los cálculos de los presupuestos mensuales de forma separada minimizando el uso de espacio en la hoja de cálculo en medida de lo posible, por esta razón, se le asignará un apartado a los presupuestos mensuales desplazados según el número de meses de retraso y recalculados según las desviaciones establecidas en las ecuaciones (6.0.1 a) y (6.0.1 b).

- *Presupuesto con retraso, extensión, variación mensual y desviación porcentual*. Con el mismo motivo explicado en el punto anterior, en este apartado se agruparán los cálculos resultantes de aplicar la variación mensual según la ecuación (6.0.2) a los resultados del apartado anterior y de forma simultánea se ejercerá la variación descrita en el punto número 5 del inicio de este capítulo.

- *Presupuesto total por subproyecto*. Es la suma total de los presupuestos mensuales de un subproyecto después de sufrir todas las variaciones establecidas, según un escenario de la simulación.

- *Presupuesto total*. Es la suma de los presupuestos totales de todos los proyectos del CAP contemplados en un escenario de simulación. A partir de este resultado se pueden tomar decisiones tales como la cantidad que debe ser asignada al conjunto de proyectos de modo que sea suficiente para no asumir un riesgo mayor al requerido, por tal motivo es el resultado más importante en el modelo completo del CAP.

- *Periodo/Fecha final*. Es el periodo mensual en el cual finalizarán todos los proyectos del CAP, o equivalentemente, el periodo en el cual finalizará el último de los proyectos del CAP, para conocer ese periodo, deben contemplarse las fechas de terminación de todos los proyectos del CAP en un escenario de la simulación y tomar la más tardía de ellas. Es evidente que este resultado se presenta de forma discreta y dado que es un resultado general, es imprescindible para el análisis y la toma de decisiones del CAP con respecto a la agilización de los proyectos y la calendarización de los mismos.

Es importante señalar que los valores alusivos a los puntos tratados en el apartado de información de salida son el resultado de la ejecución de un solo escenario de la simulación. Por lo tanto, para obtener estadísticas y resultados más útiles para la toma de decisiones es necesario ejecutar un número considerable de escenarios de simulación. Para dicho propósito, se hará uso de la hoja de cálculo Microsoft Office Excel y del software especializado en administración de riesgos Palisade Decision Tools @Risk. La construcción del modelo en hoja de cálculo se describirá de forma minuciosa en la siguiente sección.

6.2 Construcción del Modelo de Simulación

En esa sección, se explicará paso por paso la forma en que se construyó el modelo de simulación para los proyectos del CAP, así como las funciones y fórmulas de la hoja de cálculo de Excel y de @Risk requeridas para dichos fines, partiendo de la definición de las mismas funciones y el porqué de su uso individual y combinado, y desembocando en los resultados de efectuar la simulación.

6.2.1 Fórmulas y Funciones de Excel

Sin lugar a duda, las fórmulas conforman poderosas herramientas para la hoja de cálculo de Excel. Haciendo uso de varias funciones de Excel, las fórmulas permiten a los analistas generar reportes agregados, complejos motores de cálculo, construir modelos y diversidad de instrumentos más. Además, los analistas de Excel mejoran su productividad a medida que incrementa su dominio en las fórmulas y funciones de Excel.

Sin embargo, construir una buena competencia en funciones y fórmulas de Excel toma tiempo. Dado que Excel contiene más de 400 funciones, pueden invertirse meses, incluso años, aprendiendo cuales funciones son las mejores para ciertas tareas y cuales funciones pueden ser combinadas con otras funciones.

Aprovechar las funciones de Excel no sólo hace al analista más productivo, sino también le permite realizar tareas de las cuales no se tenía conocimiento de que pueden ser abordadas con las fórmulas de Excel.

A continuación, se describirán las funciones de la hoja de cálculo de Excel que se utilizaron en la construcción del modelo para el caso tratado.

- *Función SI*. La función *SI* es una de las funciones más utilizadas de Excel y permite realizar comparaciones lógicas entre un valor y un resultado esperado. Visto de una manera más sencilla, la función *SI* verifica si una proposición es Verdadera, y en caso de ser así realiza una acción específica, o de lo contrario, realiza una acción distinta. La función *SI* requiere entonces, de 3 argumentos para su empleo, y su estructura es la siguiente:

=SI(prueba_lógica, [valor_si_verdadero], [valor_si_falso]) (6.2.1)

El primer argumento es la afirmación de la cual se desea conocer su valor lógico, el segundo argumento es la acción por realizar en caso de que la prueba tenga un valor verdadero, y el tercero es la acción que se debe realizar en caso contrario.

- *Función MAX*. La función *MAX* devuelve el valor máximo del conjunto de valores que le es proporcionado. La función *MAX* tiene la siguiente sintaxis:

=MAX(número1, [número2], …) (6.2.2)

El primer argumento de esta función es obligatorio, y opcionalmente pueden añadirse de 1 a 255 argumentos, los cuales pueden ser números, nombres, matrices, fechas u otras referencias que contengan números.

- *Función Y*. Es una función lógica que ayuda a determinar si todos los argumentos que le son proporcionados son verdaderos, si esto ocurre, devuelve el valor *VERDADERO*, si por lo menos uno de los argumentos es falso, devuelve el valor *FALSO*. Esta función suele ser utilizada en combinación con la función SI en caso de ser necesario que se cumpla más de una condición incrustándose como el primer argumento de la función *SI* como se muestra en la función (6.2.3).

=SI(Y(valor_lógico1, …), [valor_si_verdadero], [valor_si_falso]) (6.2.3)

- *Función SUMA*. Esta función puede sumar valores, referencias o rangos de celda, e incluso combinaciones de las tres y su sintaxis es la siguiente:

=SUMA(número1, [número2], …) (6.2.4)

- *Función REDONDEAR*. Como el nombre lo indica, la función *REDONDEAR* redondea un número a un número de decimales especificado. La sintaxis de la función *REDONDEAR* tiene los siguientes argumentos:

=REDONDEAR(número, núm_decimales) (6.2.5)

Donde el primer argumento es el número que se desea redondear, y el segundo es el número de decimales con los que se redondeará.

- *Función DESREF*. Devuelve el valor de una celda resultante del desplazamiento en un número de columnas y en un número de filas desde una celda de referencia especificada. Los argumentos de la sintaxis de esta función son los siguientes:

  =DESREF(ref, filas, columnas, [alto], [ancho]) (6.2.6)

  El primer argumento de esta función es la referencia en la que se desea basar la desviación, y por lo tanto, debe referirse a una celda; el segundo argumento es el número de filas hacia arriba o hacia abajo al que de desea que haga referencia la celda indicada en el primer argumento, si este número es positivo, el desplazamiento será hacia abajo, y si es negativo el desplazamiento será hacia arriba; el tercer argumento es el número de columnas hacia la izquierda o hacia la derecha, al que se desea que haga referencia la celda del primer argumento; tanto el cuarto como el quinto argumento son opcionales y corresponden al alto y el ancho en número de filas y columnas que se desea que tenga la referencia devuelta.

- *Función INDICE*. La función *INDICE* devuelve un valor o la referencia a un valor desde una tabla o rango. La función *INDICE* puede ser utilizada de dos formas distintas: la forma de matriz y la forma de referencia.

  En la forma de matriz, la función *INDICE* devuelve el valor de un elemento de una tabla o matriz seleccionado por los índices de número de fila y de columna. La sintaxis para esta modalidad de la función es la siguiente:

  =INDICE(matriz, núm_fila, [núm_columna]) (6.2.7 a)

  El argumento *matriz* es un rango de celdas o una constante de matriz, el argumento *núm_fila* selecciona el número de fila de la matriz desde la cual devolverá un valor, y el argumento *núm_columna* selecciona la columna de la matriz de la cual devolverá un valor.

  La forma de referencia funciona de manera similar que la forma matricial, con la excepción de que el primer argumento es una referencia o un rango de celdas y no una matriz y cuena con un argumento opcional adicional, de esa manera, tiene la siguiente sintaxis:

  =INDICE(ref, núm_fila, [núm_columna], [núm_área]) (6.2.7 b)

  El argumento *núm_area* es utilizado para seleccionar una de las áreas del argumento de referencia cuando existe más de una.

- *Función ESTIMACION.LINEAL*. La función *ESTIMACION.LINEAL* calcula los parámetros de una regresión lineal a través del método de mínimos cuadrados encontrando la función lineal que mejor se ajusta a los datos proporcionados. Esta función también puede ser combinada con otras funciones para calcular parámetros y estadísticas de otros modelos que pese a no ser lineales pueden ser interpretados como tales, por ejemplo, funciones polinómicas, logarítmicas, exponenciales y de potencias. Es importante mencionar que esta función devuelve una matriz de valores, por lo que debe ser especificada como una fórmula de matriz. La sintaxis de la función es la siguiente:

=ESTIMACION.LINEAL(conocido_y, [conocido_x], [constante], [estadística]) (6.2.8)

El argumento *conocido_y* es el conjunto o rango de valores conocidos de la variable dependiente, el argumento *conocido_x* es el conjunto o rango de valores conocidos de la variable o variables independientes, el argumento *constante* es opcional y sirve para forzar el modelo a uno cuyo valor del intercepto es igual a cero, si es omitido este argumento o le es asignado el valor *VERDADERO*, la estimación se realiza de manera normal, el argumento *estadística* sirve para indicar si se desean obtener estadísticas adicionales como los errores estándar, el coeficiente de determinación, etc.

Las funciones anteriores son las que serán útiles para la construcción del modelo y con las que cuenta por defecto la paquetería de Microsoft Office 2016. Sin embargo, también es necesario utilizar algunas proporcionadas por el software @Risk.

6.2.2 Fórmulas y Funciones de @Risk

Cuando un proyecto es abordado en Excel, las herramientas de modelación de @Risk disponibles para Excel pueden usarse para construir un modelo de riesgo asociado al proyecto. Una muestra de lo anterior es la capacidad de asignar una distribución de probabilidad a una celda que funcione como un recurso útil para el modelo.

La definición de distribuciones de probabilidad en la hoja de cálculo puede realizarse por dos caminos, uno de ellos es a través del intuitivo menú que posee el software y otro es insertando las fórmulas directamente en las celdas.

Si se tiene una mínima experiencia utilizando el software, la forma más práctica para definir los modelos de probabilidad deseados en las celdas de la hoja de cálculo es insertando la función asociada, tal como se hace con una función de Excel.

A continuación, se describirán las funciones de @Risk utilizadas para la construcción del modelo del CAP.

- *Función RiskBernoulli*. Esta función especifica una distribución Bernoulli con probabilidad "p" de éxito en cada iteración. El $(p \times 100)\%$ de las veces, la función generará el valor 1, y su sintaxis es la siguiente:

=RiskBernoulli(P) (6.2.9)

En este caso, el argumento *P* puede ser una constante con valor entre 0 y 1 o bien, una celda o referencia cuyo valor se encuentre entre 0 y 1.

- *Función RiskPert*. Especifica una distribución PERT con valores *mínimo y máximo* especificados. El parámetro de forma se calcula a partir del valor *más probable*. La distribución PERT puede ser utilizada como una distribución muy práctica y fácil de entender. La distribución PERT puede ser considerada, en general, como una distribución superior a la Triangular cuando los parámetros resultan en una distribución sesgada, ya que su forma suavizada pone más énfasis en la dirección del sesgo. La sintaxis de esta función es bastante sencilla, basta con introducirse los valores de los parámetros, es decir, el mínimo, la moda y el máximo, en ese orden, tal como se muestra a continuación:

=RiskPert(mínimo, más probable, máximo) (6.2.10)

- *Función RiskOutput*. La función *RiskOutput* se utiliza para identificar las celdas de salida seleccionadas en la hoja de cálculo. Esta función tiene tres argumentos opcionales, como se muestra a continuación:

=RiskOutput("nombre de la celda de salida", "nombre de rango de salida", núm. de elemento en el rango") (6.2.14)

La función *RiskOutput* puede introducirse añadiéndola a la fórmula existente de la celda que se va a seleccionar como salida de simulación, por ejemplo, si una celda contiene la función *=SUMA(1,2)* al elegirse como salida pasará a ser *=RiskOutput()+SUMA(1,2)* en caso de no introducirse argumentos en la función.

- *Función RiskMean*. Genera la media de la distribución simulada para una referencia de celda. La sintaxis de la función es la siguiente:

=RiskMean(ref_celda o nombre de salida/entrada) (6.2.11)

El argumento para esta función también puede ser el nombre de una salida que se ha definido previamente.

- *Función RiskVariance*. Genera la varianza de la distribución simulada para una referencia de celda. Esta función trabaja de la misma forma que *RiskMean* como puede observarse en su sintaxis:

$$=\mathit{Riskvariance(ref\_celda\ o\ nombre\ de\ salida/entrada)} \qquad (6.2.12)$$

- *Función RiskPercentile*. General el valor del percentil deseado, el cual puede ser una constante o una referencia de celda, de la distribución simulada para una referencia de celda o el nombre de una salida. La sintaxis de esta función se muestra a continuación:

$$=\mathit{RiskPercentile(ref\_celda\ o\ nombre\ de\ salida/entrada,\ percentil\ )} \qquad (6.2.13)$$

Las funciones anteriores pueden añadirse navegando en el menú de @Risk en donde se encuentra la miscelánea de funciones, esto se puede observar en la siguiente ilustración:

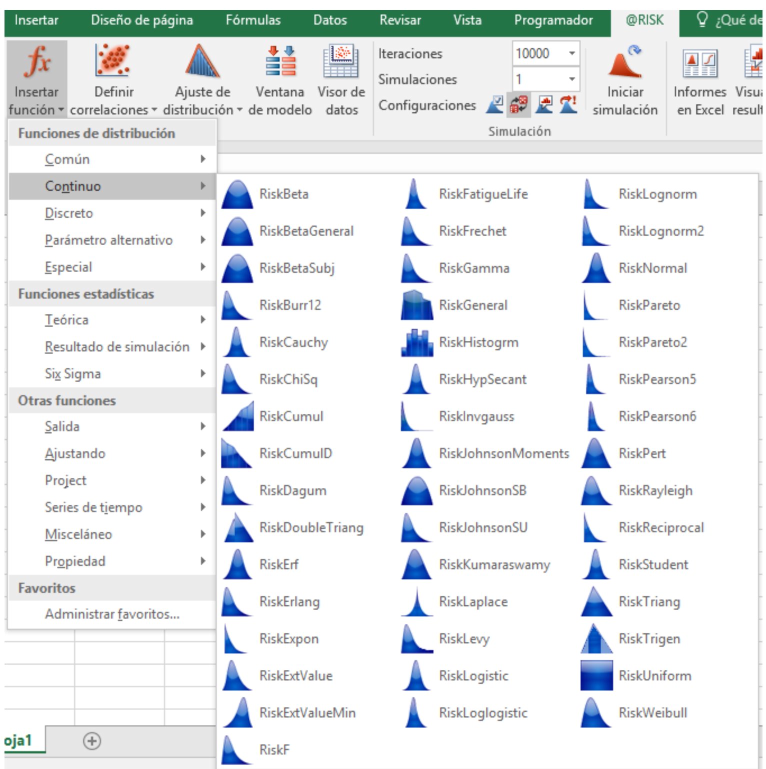

Figura 6.3: Miscelánea de Funciones de @Risk. Fuente: Elaboración propia

Las herramientas explicadas anteriormente son suficientes para construir y explicar de manera detallada el modelo de los proyectos del CAP en la hoja de cálculo de Excel combinando las funciones de ambas paqueterías con los conceptos probabilísticos y estadísticos desarrollados en los capítulos anteriores.

6.2.3 Estructuración del Modelo

En la Tabla 6.4 puede apreciarse que los 31 subproyectos del CAP se agruparon en 17 proyectos distintos, sin embargo, para los fines que competen a este modelo de simulación, esta forma de agrupar los subproyectos no interviene en el planteamiento de las fórmulas o en los cálculos requeridos, es por ello que una forma de dar una estructura práctica al modelo es utilizando una hoja de cálculo para cada uno de los proyectos que involucren con más de un subproyecto, es decir, los proyectos CAP1, CAP2, CAP3 y CAP4, pero, uno de los objetivos de este proyecto es optimizar el uso de recursos computacionales, por esta razón, es conveniente agrupar los proyectos cuyo número de subproyectos es 1, formando 3 grupos: el primero conteniendo los 5 subproyectos desde CAP5 hasta CAP9, el segundo también con 5 subproyectos, siendo estos del CAP10 al CAP14, y el tercer grupo con los 3 subproyectos restantes, del CAP15 al CAP17, teniendo así 7 grupos de subproyectos y por consiguiente 7 hojas de cálculo con información de los subproyectos.

Cada una de las 7 hojas del libro contiene en las 2 primeras columnas los periodos mensuales en formato de fecha y con el número correspondiente en la primera y segunda columna de manera respectiva, ubicándose la fecha 30/04/2014 en la forma "*abr-14*" en la celda *A9* y el número 1 que corresponde a dicho periodo en la celda *B9* bajo las etiquetas de columna *Periodo* y # *Periodo* que se encuentran en las celdas *A8* y *B8* respectivamente. Tal como en el periodo de abril de 2014 los siguientes periodos se ubican en las celdas correspondientes ubicándose "*may-14*" en la celda *A10* y el número 2 en la celda *B10*, y de manera sucesiva hasta el periodo comprendido en el mes de abril del año 2027 en el formato "*abr-27*" y su número de periodo 157 ubicados en las celdas *A165* y *B165* en ese orden. Como se mencionó anteriormente, a razón de contemplar este rango de periodos se halla en el hecho de que ningún periodo comenzó antes del mes de abril del año 2014 y pese a las posibles demoras el periodo de abril del año 2027 es lo suficientemente lejano para contemplar todos los periodos en los que pueda existir actividad de los proyectos. En la siguiente ilustración se muestra lo descrito anteriormente:

	A	B
8	Periodo	# Periodo
9	abr-14	1
10	may-14	2
11	jun-14	3

Figura 6.4: Estructuración de los periodos mensuales. Fuente: Elaboración propia

Independientemente de las columnas utilizadas para describir los periodos mensuales, cada subproyecto requiere de por lo menos 4 columnas para la presentación de sus datos, por lo cual, la información de esos subproyectos, incluyendo las etiquetas de la misma, debe ser distribuida de manera óptima en ese espacio. Para los anteriores propósitos, primero debe ser ingresada y distribuida la información de entrada, ésta se colocó la parte posterior de las hojas en un apartado que consta de 4 columnas y 6 filas distribuida de la siguiente manera:

	C	D	E	F
2		# Subproyecto:	1	CAP1-P1
3	Fecha de Inicio	31/05/2014	Fecha de Fin	30/11/2016
4	Periodo de referencia	27	Retraso	0
5	Periodo inicial estimado	2	Extensión	1
6	Periodo final estimado	31	¿Se Inicial el Proyecto? (0=No, 1=Sí)	1
7			Porcentaje Ejercido	89.17%

Figura 6.5: Ejemplo de la distribución de la información de entrada de un subproyecto en la hoja de Excel. Fuente: Elaboración propia

En la figura 6.5 se observan las etiquetas de algunos datos de entrada en la celda a la izquierda de los datos que le corresponden, de igual manera, se encuentran las etiquetas de la Fecha de Inicio y la Fecha de Fin, pero, dado que son datos de salida, por el momento se mostrará vacíos dichos campos. En la ilustración, se muestran las celdas que corresponden al subproyecto 1 del proyecto 1, o bien, el proyecto *CAP1-P1*, en este caso, se utilizaron las columnas *C, D, E* y *F*, es decir, de la tercera a la sexta columna de la hoja de cálculo, así que, para cada subproyecto debe utilizarse el rango de columnas según la relación (*4i-1:4i+2*), donde *i* es el número de subproyecto contenido en la hoja de cálculo, por ejemplo, para el subproyecto número 4, debe utilizarse el rango de la 15va a la 18va columna, es decir, de la columna *O* a la columna *R* , en el rango de la fila número 2 a la fila número 7 de la hoja que le corresponda. En la celda ubicada en la tercera columna del rango correspondiente y la fila número 2 de la hoja de cálculo se ubica el número de subproyecto contenido en la hoja de cálculo, y a su derecha la etiqueta del mismo, en este caso, el número 1 y la etiqueta *CAP1-P1*; en la celda ubicada en la segunda columna del rango correspondiente y la cuarta fila de la hoja, en este caso la celda *D4*, se encuentra el periodo de referencia en el cual se perdió la continuidad de la base de datos, y para todos los casos será la constante 27, que corresponde al periodo de junio de 2016, tal como se planteó anteriormente; una celda abajo, en la quinta fila de la hoja y la segunda columna, del rango correspondiente al subproyecto (*D5* para este caso), se ubica el periodo de inicio estimado del proyecto, es decir, la fecha contemplada para el inicio del proyecto en caso de no existir retrasos, en este caso el número 2 que corresponde al mes de mayo de 2014; en la celda ubicada en la fila posterior (*D6* para este caso), está colocado el periodo de finalización estimado para el subproyecto, en este caso, el número 31, el cual está asociado al mes de octubre del año 2016, fecha contemplada para terminar el proyecto en caso de no existir retraso en la fecha de inicio u otra demora durante el desarrollo del proyecto. Nótese que las celdas descritas anteriormente pueden ser llenadas únicamente con constantes, sin necesidad de utilizar fórmulas, esto se debe a que contienen información puramente de entrada, a diferencia de ello, las filas restantes involucran información de entrada y valores resultantes de la ejecución de por lo menos una iteración de la simulación. En la celda de la cuarta columna del rango correspondiente al subproyecto y la cuarta fila de la hoja (*F4* para este caso), se encuentra el número de meses de retraso en la fecha de inicio del proyecto, que como se justificó en la sección 6.1 está asociado a un valor aleatorio generado de acuerdo a una distribución PERT redondeado al entero más cercano, en este caso le proyecto *CAP1-P1* ya había iniciado antes de la fecha de referencia (en el mes de mayo de 2014), por lo cual, no hay posibilidad alguna de que el número de meses de retraso sea mayor a 0, y por consiguiente, la fórmula asociada a esta variación es *=REDONDEAR(RiskPert(0,0,0),0)* la cual siempre generará el valor 0. En la celda de la cuarta columna del rango correspondiente al subproyecto y la quinta fila de la hoja (en este caso *F5*) se ubica el número de meses de extensión del proyecto que también está asociado a una distribución PERT redondeada al entero más cercano y que en el caso del proyecto *CAP1-P1* tiene un valor mínimo de 0, un valor más probable de 0 y un máximo de 3, por lo que la fórmula asociada a esta variación es *=REDONDEAR(RiskPert(0,0,3),0)*; en la misma columna, y en la fila posterior, se halla la variable que determina si se ejecutará o no el proyecto, y como se mencionó anteriormente, al existir solamente 2 posibilidades, la distribución asociada a este evento es una distribución Bernoulli, y dado que el proyecto *CAP1-P1* ya tuvo inicio, su probabilidad de ejecución es 1, así que la fórmula para la celda *F6* es *=Riskbernoulli(1)* la cual generará el valor 1 en el 100% de las iteraciones efectuadas. En la intersección de la cuarta columna del rango de cada proyecto y la fila número 7 de la hoja correspondiente se encuentra la desviación porcentual con respecto a la totalidad de presupuesto del proyecto, cuya definición se dio al inicio de este capítulo, mismo en el que se determinó que la mejor distribución para representarlo es la distribución PERT sin modificación alguna, y al caso del proyecto *CAP1-P1* se le asocian los porcentajes mínimo, más probable y máximo de 75%, 90% y 100%, así que, para efectos prácticos, la fórmula que le corresponde es *=RiskPert(0.75, 0.9, 1)*. Finalmente, en las intersecciones de la segunda y cuarta columna del rango perteneciente a cada proyecto y la tercera fila de la hoja de cálculo, se hallan las fechas de inicio y terminación del proyecto de manera respectiva. Estas fechas dependen de las fechas de inicio y terminación estimadas y del número de meses de retraso y extensión, tal como se aprecia en las ecuaciones (6.1.1) y (6.1.2), aunque resulta más práctico visualizar los momentos de inicio y terminación de los proyectos en formato de fecha en lugar del número de periodo, es más fácil realizar operaciones con el número de periodo asociado a una fecha que con la fecha misma. Una de las ventajas que proporciona la estructura que se le dio al modelo dentro de la hoja de cálculo radica en la facilidad de asociar una fecha con su número de periodo, esta ventaja puede ser aprovechada realizando desplazamientos en las filas de la hoja, esto se logra por medio de la función *DESREF* descrita en la ecuación (6.2.6), si se toma como referencia fija a la celda *A8* que es la celda anterior a las de las fechas utilizadas para el modelo, y los desplazamientos son el número de periodo de inicio y finalización del proyecto para el número de filas de desplazamiento y dado que las fechas se encuentran en la misma columna, se pueden obtener las fechas exactas de inicio y terminación del proyecto. Por ejemplo, para obtener la fecha de inicio del proyecto *CAP1-P1* ubicada en la celda *D3* se utiliza la fórmula *=DESREF(A8, D5+F4, 0)*, donde *A8* se mantiene fija para poder ser emulada para el resto de los proyectos, *D5+F4* es el periodo de inicio asociado a la variación simulada, y 0 es el número de columnas de desplazamiento desde la celda *A8*. Por último, la fecha de terminación puede obtenerse de la misma forma, con la excepción de que el número de filas de desplazamiento está dado por *D6+F4+F5* que es el número periodo de terminación determinado por la ecuación (6.1.2), resultando así, en la fórmula *=DESREF(A8, D6+F4+F5)*, y al igual que la fórmula para la fecha de inicio, ésta puede ser emulada por cualquier otro subproyecto de la hoja, sin hacer mayor tarea a copiar el contenido de la celda en el lugar correspondiente.

Como puede observarse en el ejemplo de la figura 6.5, en las celdas con fondo azul, que son donde se incluyen las entradas del modelo, se han generado valores arbitrarios pese a no haberse efectuado la simulación, por lo general, las fórmulas de @Risk imponen por defecto el valor de la media de la distribución definida en la celda, siempre que éste se encuentre en el conjunto de valores de la distribución, por ejemplo, en el caso del número de meses de extensión se tienen los parámetros 0, 0 y 3, y la media asociada a esa distribución es igual a 0.5, pero, como se le ha incrustado la función *REDONDEAR* muestra el valor 1, y en el caso de las distribuciones discretas, muestra el entero más cercano a la media. Para generar valores distintos, debe activarse la opción *Recálculo aleatorio/estático* de @Risk.

En la parte inferior la sección descrita en los párrafos anteriores, se hallan los apartados con los títulos *Periodo proyectado*, *Presupuesto original*, *Presupuesto con retraso y extensión* y *Presupuesto con retraso, extensión, variación mensual y desviación porcentual*, mismos que a excepción del *Periodo proyectado* se encuentran definidos en la sección 6.1 y se formularan a continuación. En el apartado *Periodo proyectado* se halla el número de periodo futuro para el proyecto, es decir, el número de periodo en que se realizará una actividad del proyecto partiendo de la fecha de referencia, siendo ésta la asociada al número 1. Por ejemplo, si un proyecto inicia en el mes de mayo del año del 2017, su fecha de inicio tiene asociado el periodo proyectado número 1, dado que ésta es posterior a la fecha de referencia. En resumen, sólo los periodos de un proyecto con un presupuesto asignado y que son iguales o posteriores al mes de junio de 2016 tienen un número de periodo proyectado, correspondiendo el 1 al más temprano de ellos. En la siguiente figura se observa la distribución de los apartados mencionados en la hoja de cálculo:

	A	B	C	D	E	F
8	Periodo	# Periodo	Periodo Proyectado	Presupuesto Original	Presupuesto con retraso y extensión	Presupuesto con retraso, extensión, variación mensual y desviación porcentual
9	abr-14	1		£0.00	£0.00	£0.00
10	may-14	2		£759,627.29	£759,627.29	£759,627.29
11	jun-14	3		£759,627.29	£759,627.29	£759,627.29
12	jul-14	4		£759,627.29	£759,627.29	£759,627.29
13	ago-14	5		£780,221.87	£780,221.87	£780,221.87

Figura 6.7: Estructura del modelo en la hoja de cálculo. Fuente: Elaboración propia

En la imagen anterior, las celdas de la columna *C* se encuentran vacías debido a que los periodos mostrados son anteriores al mes de junio de 2016, por esa misma razón, no hay diferencia entre los valores de las 3 columnas siguientes. En la segunda de las 4 columnas utilizadas por cada proyecto se encuentran los presupuestos mensuales ordenados según el periodo que les corresponde, estos valores son los propuestos inicialmente sin sufrir ninguna variación. En la segunda columna, se ubican los valores de los presupuestos después de contemplarse los meses de retraso en la fecha de inicio y la extensión en el número de meses requeridos, esta sección debe tratarse con cautela debido al sinnúmero de posibilidades que existen con respecto a las fechas de ejecución del proyecto. La fórmula correspondiente a la celda *E2* es la que se muestra en seguida:

```
= SI(Y($B9<27,D$5+F$4<27),DESREF(E9,0,-
1), SI(Y($B9>=MAX(27,D$5+F$4),$B9<=D$6+F$4),DESREF(E9,-
1*F$4,-1)*((D$6-D$5+1-CONTAR.SI(D$9:D$34,">0"))/(D$6-D$5+1-
CONTAR.SI(D$9:D$34,">0")+F$5)), SI(Y($B9>D$6+F$4,$B9<=D$6+F$
4+F$5),(D$166-SUMA(E$9:DESREF(E$8,D$6+F$4,0)))/F$5,"")))
```
(6.2.14)

Puede observarse que el texto de la formula anterior tiene 3 colores distintos, esos 3 colores representan cada una de las funciones *SI* y su respectiva condición, puede notarse que la función condicional de color verde se encuentra anidada en la función condicional en color azul a manera del argumento *valor_si_verdadero* y a su vez ésta se encuentra anidada de la misma forma en la condicional de color rojo. La primera función *SI* (en color rojo) se encarga de verificar 2 condiciones: que el periodo correspondiente al presupuesto tratado en la fórmula sea menor a la fecha de referencia, y que la fecha de inicio del proyecto sea anterior a la fecha de referencia, si estás 2 condiciones se cumplen, la función devolverá el valor de la celda que resulte del desplazamiento de la celda actual en 0 filas y -1 columnas, es decir, el valor de la celda que se encuentra a la izquierda, esta condición garantiza que cuando un proyecto se inició en una fecha anterior a junio de 2016, los valores de los presupuestos de los meses transcurridos se conserven, debido a que no es posible que sufran alguna alteración al haberse gastado dichos montos. En caso de no cumplirse alguna de las 2 condiciones, los valores pasarán por un nuevo filtro proporcionado por la función *SI* escrita en color azul, esta función verificará que se cumplan 2 nuevas condiciones: que el periodo al que corresponde el presupuesto de la celda actual sea cuanto más temprano el máximo entre el periodo de referencia y la fecha de inicio del proyecto y no rebase la fecha de terminación con retraso y sin extensión, en caso de cumplirse ambas condiciones la función devolverá el valor de la celda que resulte de desplazarse un número de filas hacia arriba equivalente al número de retrasos en la fecha de inicio y una columna hacia la izquierda multiplicado por el cociente del número de meses restantes del proyecto en caso de no existir extensión sobre el número de meses restantes en caso de existir extensión, esto es con base en lo planteado por la fórmula (6.0.1 a); de no cumplirse alguna de las condiciones, la fórmula establecerá una última función con 2 nuevas condiciones: que el periodo correspondiente al presupuesto de la celda actual rebase la fecha de terminación en caso de no existiera alguna extensión sin rebasar la fecha de terminación del proyecto habiendo extensión, si se cumplen las 2 condiciones, la fórmula devolverá el cociente de la resta del monto total del proyecto menos el monto total de los meses transcurridos sobre el número de meses de extensión, este cálculo es de acuerdo con la fórmula (6.0.1 b), esta condición asignará los montos pertenecientes a los periodos que sólo se contemplarán en caso de existir una prolongación en el número de meses requeridos para el proyecto, si por lo menos una de estas condiciones no es satisfecha, la fórmula procederá a dejar vacía la celda. En resumen, las 3 funciones condicionales propuestas anteriormente tienen la tarea de verificar la situación temporal de cada periodo y con base en eso asignarles el presupuesto que les corresponde habiendo contemplado las demoras en el inicio y terminación del proyecto. Además, la forma en que está estructurada la fórmula permite su generalización para cualquier celda del mismo apartado para cualquiera de los proyectos.

La siguiente formula corresponde a la celda *C9*, y en general, a la primera columna del rango de un proyecto, y es esta fórmula la que define el número de periodo proyectado para el punto de partida de las observaciones de la variable independiente de la regresión polinomial asociada a la variación mensual de los presupuestos después de las demoras, o en su defecto, no asigna valor alguno en caso de no corresponderle:

$$=SI(Y(E9<>"",E9<>0),\ SI(\$B9>=MAX(D\$4,D\$5+F\$4),\ SI(\$B9=MAX(D\$4,D\$5+F\$4),1,C8+1),""),"") \qquad (6.2.15)$$

A diferencia de la fórmula (6.2.14), esta fórmula anida las funciones condicionales a partir del primer argumento (*prueba_lógica*) de la función *SI*, de esa forma, se verificarán en orden las condiciones y posteriormente se realizarán las acciones en caso de no satisfacerse. La primera condición escrita en rojo, analiza si la celda correspondiente al presupuesto con retraso y extensión tiene algún valor y si éste es distinto de 0, de ser así, generará un nuevo filtro que determinará si el periodo que concierne la celda que contiene la fórmula es mayor o igual al periodo más tardío entre el mes de junio de 2016 y la fecha de inicio del proyecto contemplando el retraso, dado que el mayor de los periodos comparados es al que se le asigna el número 1 de los periodos proyectados, y la condición anterior garantiza que cada uno de los periodos proyectados es posterior o igual a éste, sólo queda generar un filtro más que determine la asignación de los valores, en este caso la condicional en color verde se encara de verificar que una vez que se han cubierto los requisitos anteriores cuál de ellos es el menor, y es al cual le asignará el valor 1, de ser posterior a él, le asignará el valor del periodo proyectado de la celda que se encuentra una fila arriba aumentado en 1, por ejemplo, si determina que al periodo 27 le corresponde el valor 1, le asignará el valor 2 al periodo 28, y así sucesivamente hasta el periodo de finalización del proyecto. Al igual que las formulas anteriores, esta se encuentra generalizada y basta con copiar el contenido de la celda en la celda adecuada. Para concluir la parte elemental del modelo y proceder a la ejecución de la simulación, se requiere de una fórmula lo suficientemente robusta para atender a las variaciones establecidas anteriormente y las posibles desviaciones en el presupuesto determinadas por los puntos 4 y 5 establecidos al inicio de este capítulo. Un ejemplo generalizable de esa fórmula es el siguiente:

```
= SI(E9<>"", SI(Y($B9>=MAX(D$4,D$5+F$4),$B9<=D$6+F$4+F$5), ((F$7*D$166-
SUMA(E$9:E$34))/SUMA(E$35:E$165)) * (INDICE(ESTIMACION.LINEAL(DESREF($B$9,M
AX(D$4,D$5+F$4)-$B$9,4*E$2-1):DESREF($B$9,D$6+F$4+F$5-$B$9,4*E$2-
1),(DESREF($B$9,MAX(D$4,D$5+F$4)-$B$9,4*E$2-3):DESREF($B$9,D$6+F$4+F$5-
$B$9,4*E$2-3))^{1,2,3,4}),1,1) *(DESREF(F9,0,-
3)^4)+ INDICE(ESTIMACION.LINEAL(DESREF($B$9,MAX(D$4,D$5+F$4)-$B$9,4*E$2-
1):DESREF($B$9,D$6+F$4+F$5-$B$9,4*E$2-1),(DESREF($B$9,MAX(D$4,D$5+F$4)-
$B$9,4*E$2-3):DESREF($B$9,D$6+F$4+F$5-$B$9,4*E$2-
3))^{1,2,3,4}),1,2) *(DESREF(F9,0,-
3)^3)+ INDICE(ESTIMACION.LINEAL(DESREF($B$9,MAX(D$4,D$5+F$4)-$B$9,4*E$2-
1):DESREF($B$9,D$6+F$4+F$5-$B$9,4*E$2-1),(DESREF($B$9,MAX(D$4,D$5+F$4)-
$B$9,4*E$2-3):DESREF($B$9,D$6+F$4+F$5-$B$9,4*E$2-
3))^{1,2,3,4}),1,3) *(DESREF(F9,0,-
3)^2)+ INDICE(ESTIMACION.LINEAL(DESREF($B$9,MAX(D$4,D$5+F$4)-$B$9,4*E$2-
1):DESREF($B$9,D$6+F$4+F$5-$B$9,4*E$2-1),(DESREF($B$9,MAX(D$4,D$5+F$4)-
$B$9,4*E$2-3):DESREF($B$9,D$6+F$4+F$5-$B$9,4*E$2-3))^{1,2,3,4}),1,4) *DESREF(F9,0,-
3)+ INDICE(ESTIMACION.LINEAL(DESREF($B$9,MAX(D$4,D$5+F$4)-$B$9,4*E$2-
1):DESREF($B$9,D$6+F$4+F$5-$B$9,4*E$2-1),(DESREF($B$9,MAX(D$4,D$5+F$4)-
$B$9,4*E$2-3):DESREF($B$9,D$6+F$4+F$5-$B$9,4*E$2-3))^{1,2,3,4}),1,5) ) ,E9) ,"")
```

(6.2.16)

La fórmula (6.2.16) contiene únicamente 2 funciones condicionales, donde la labor de la primera de ellas es verificar que la celda del presupuesto con retraso y extensión asociada al periodo analizado no se encuentre vacía, es decir, que la acción indicada en el siguiente argumento se realice sólo si se tiene un presupuesto asignado, y de no ser así, la celda que contiene la fórmula se quedará vacía; la segunda función condicional está escrita en color azul y tiene la tarea de clasificar el periodo subyacente a la celda actual en periodos concluidos o periodos próximos, si determina que se trata de un periodo concluido, arrojará el valor del presupuesto original, contrario a ello, si se determina que el periodo asociado a la celda que contiene la fórmula es un periodo futuro a concluirse se llevaran a cabo las operaciones de las funciones de la fórmula escritas en los colores anaranjado, verde y morado cuyas explicaciones se darán en seguida.

Las regiones de la fórmula (6.2.16) escritas en color verde son probablemente las más importantes de este apartado, debido a que se encargan de hallar los estimadores de los parámetros cuadrados de la regresión polinomial asociada al proyecto analizado por el método de mínimos cuadrados. La estimación de dichos parámetros se logra a través del uso combinado de las funciones *ESTIMACION.LINEAL* e *INDICE*, la primera anidada dentro de la segunda, en la fórmula se halla 5 veces esta combinación, siendo el tercer argumento de la función *INDICE* la única diferencia entre ellas, el cual es igual a 1 para la primera de ellas, 2 para la segunda y de forma sucesiva hasta la quinta, en la fórmula (6.2.7 a) se mencionó que el tercer argumento de la función *INDICE* es el número de columna en el que se ubica la referencia de la matriz que se desea presentar, y la fórmula *ESTIMACION.LINEAL* devuelve los parámetros estimados de una función lineal en un vector horizontal, así que esta combinación devolverá los 5 estimadores deseados en el orden que se muestra en la ecuación (6.0.2), en cada una de estas 5 funciones se introducirán los valores de la variable dependiente en el primer argumento, cuyo rango comenzará en la celda de la tercera columna del rango perteneciente al subproyecto y la primer fila a la que le corresponda un periodo proyectado y terminará en la celda ubicada en la misma columna y en la última fila con un periodo proyectado asignado, esto se hará por medio de un desplazamiento desde la columna *B* a un número de columnas determinado por la relación *4k-1*, dónde *k* es el número de subproyecto dentro de la hoja. Las celdas que contendrán los valores de la variable independiente se obtendrán de forma similar que con las de la variable dependiente, con la excepción de que el desplazamiento en el número de columnas será determinado por *4k-3*, lo cual siempre hará referencia a los valores de la primera columna del rango perteneciente al proyecto, misma que contiene los valores de los periodos proyectados. Pero, al tratarse de una regresión polinomial y no lineal deben atenderse las potencias de la variable independiente, por lo cual se agregó la operación *"^{1,2,3,4}"* al final de la referencia del rango de valores de dicha variable, este comando elevará los valores del periodo proyectado a las potencias 1, 2, 3 y 4 generando 4 variables a partir de una: $t,\ t^2, t^3$ y t^4.

Después de estimar los parámetros de la regresión polinomial, las partes de la fórmula se encargarán de multiplicarlos por las potencias de la variable t que les correspondan, esto se realiza por medio de un desplazamiento de 3 columnas hacia la izquierda y ninguna fila, de esa forma, los valores que elevará serán los del periodo proyectado que corresponde a la celda que contenga la fórmula.

Los cálculos descritos anteriormente se efectúan con el fin de generar los valores resultantes de aplicar las desviaciones en los presupuestos mensuales, sin embargo, queda una variación más que debe contemplarse en la fórmula: la desviación porcentual en el total del presupuesto del proyecto. Para esta última variación es necesario tomar en cuenta que, dado que pueden existir periodos transcurridos en un proyecto, no es posible aplicar el porcentaje generado por la distribución *PERT* de igual manera a todos los montos mensuales del proyecto, sino que debe hallarse un porcentaje aplicable a los montos de los periodos posteriores a la fecha de referencia de tal forma que la suma de los montos de la última columna sea igual al monto presupuestado originalmente multiplicado por el porcentaje generado por la distribución *PERT*. Dicho porcentaje puede calcularse de acuerdo con la siguiente ecuación:

$$\frac{PE * PT - \sum_{i=1}^{26} P_i}{\sum_{i=26}^{157} P_i} \tag{6.2.17}$$

donde PE es el porcentaje ejercido sobre el monto total del proyecto, PT es el monto total del proyecto presupuestado inicialmente, y P_i son los montos mensuales del proyecto en el periodo i contemplando retrasos en la fecha de inicio y prolongación en el número de meses requeridos. La fórmula calcula el monto total del proyecto después de sufrir todas las variaciones posibles, posteriormente a ese monto le resta la suma de los montos de los periodos transcurridos y finalmente divide ese resultado por la suma de los montos de los periodos no transcurridos. Si se multiplican los montos de los periodos no transcurridos, el monto total se equilibrará de tal forma que será igual al establecido inicialmente multiplicado por la variación porcentual. El cociente de la ecuación (6.2.17) se calcula en términos de las celdas en la región anaranjada de la fórmula (6.2.16).

Con esta última fórmula se han cubierto las variaciones correspondientes a los puntos 2, 3, 4 y 5 descritas al inicio del capítulo. Finalmente, en la celda *F166* (o la intersección de la fila 166 y la cuarta columna del rango perteneciente al proyecto) se coloca la suma de los montos mensuales del proyecto después de las variaciones, y para contemplar el hecho de que el proyecto pudiera no llevarse a cabo, basta con multiplicar dicho monto por el valor generado por la distribución *Bernoulli,* (tal como en este caso, donde la celda *F166* contiene la fórmula: *=SUMA(F9:F165)*F6*) , de esa forma, asignará el monto total del proyecto después de las variaciones en caso de ejecutarse el proyecto y asignará un presupuesto de 0 libras en caso contrario.

Los procedimientos mostrados anteriormente deben en gran parte su complejidad a la facilidad que tienen para generalizar su uso en cualquier celda de la misma columna, ya sea en el mismo proyecto o en un distinto, y a la capacidad de observar en tiempo real las variaciones y movimientos generados por la ejecución de las iteraciones de la simulación cuando se encuentre activado el recálculo aleatorio. En la siguiente figura se muestra amanera de ejemplo el prorrateo para el proyecto *CAP2-P2* en 2 iteraciones distintas cuyos periodos de extensión son de 2 meses para el primer caso y de 5 meses para el segundo:

	A	B	G	H	I	J
1						
2				# Subproyecto:	2	CAP2-P2
3			Fecha de Inicio	30/06/2015	Fecha de Fin	30/01/2017
4			Periodo de referencia	27	Retraso	0
5			Periodo inicial estimado	15	Extensión	2
6			Periodo final estimado	32	¿Se Inicial el Proyecto? (0=No, 1=Sí)	1
7					Porcentaje Ejercido	92.23%
8	Periodo	Periodo	Periodo Proyectado	Presupuesto Original	Presupuesto con retraso y extensión	Presupuesto con retraso, extensión, variación mensual y desviación
23	jun-15	15		£4,157,040.96	£4,157,040.96	£4,157,040.96
24	jul-15	16		£865,218.19	£865,218.19	£865,218.19
25	ago-15	17		£1,115,060.76	£1,115,060.76	£1,115,060.76
26	sep-15	18		£716,552.50	£716,552.50	£716,552.50
27	oct-15	19		£953,839.27	£953,839.27	£953,839.27
28	nov-15	20		£1,330,166.20	£1,330,166.20	£1,330,166.20
29	dic-15	21		£1,713,970.66	£1,713,970.66	£1,713,970.66
30	ene-16	22		£888,891.67	£888,891.67	£888,891.67
31	mar-16	23		£814,129.28	£814,129.28	£814,129.28
32	mar-16	24		£759,977.32	£759,977.32	£759,977.32
33	abr-16	25		£807,800.38	£807,800.38	£807,800.38
34	may-16	26		£659,269.72	£659,269.72	£659,269.72
35	jun-16	27	1	£409,375.40	£307,031.55	£255,796.13
36	jul-16	28	2	£1,371,027.45	£1,028,270.59	£557,297.64
37	ago-16	29	3	£322,982.67	£242,237.01	£553,319.39
38	sep-16	30	4	£1,788,866.97	£1,341,650.23	£484,545.76
39	oct-16	31	5	£106,901.94	£80,176.46	£480,166.59
40	nov-16	32	6	£1,568,139.35	£1,176,104.51	£557,877.17
41	dic-16	33	7	£0.00	£695,911.72	£623,878.23
42	ene-17	34	8	£0.00	£695,911.72	£472,875.99
43	mar-17	35		£0.00		
44	mar-17	36		£0.00		
45	abr-17	37		£0.00		
46	may-17	38		£0.00		

	A	B	G	H	I	J
1						
2				# Subproyecto:	2	CAP2-P2
3			Fecha de Inicio	30/06/2015	Fecha de Fin	30/04/2017
4			Periodo de referencia	27	Retraso	0
5			Periodo inicial estimado	15	Extensión	5
6			Periodo final estimado	32	¿Se Inicial el Proyecto? (0=No, 1=Sí)	1
7					Porcentaje Ejercido	87.95%
8	Periodo	Periodo	Periodo Proyectado	Presupuesto Original	Presupuesto con retraso y extensión	Presupuesto con retraso, extensión, variación mensual y desviación
23	jun-15	15		£4,157,040.96	£4,157,040.96	£4,157,040.96
24	jul-15	16		£865,218.19	£865,218.19	£865,218.19
25	ago-15	17		£1,115,060.76	£1,115,060.76	£1,115,060.76
26	sep-15	18		£716,552.50	£716,552.50	£716,552.50
27	oct-15	19		£953,839.27	£953,839.27	£953,839.27
28	nov-15	20		£1,330,166.20	£1,330,166.20	£1,330,166.20
29	dic-15	21		£1,713,970.66	£1,713,970.66	£1,713,970.66
30	ene-16	22		£888,891.67	£888,891.67	£888,891.67
31	mar-16	23		£814,129.28	£814,129.28	£814,129.28
32	mar-16	24		£759,977.32	£759,977.32	£759,977.32
33	abr-16	25		£807,800.38	£807,800.38	£807,800.38
34	may-16	26		£659,269.72	£659,269.72	£659,269.72
35	jun-16	27	1	£409,375.40	£223,295.67	£168,865.59
36	jul-16	28	2	£1,371,027.45	£747,833.15	£266,627.54
37	ago-16	29	3	£322,982.67	£176,172.37	£308,118.08
38	sep-16	30	4	£1,788,866.97	£975,745.62	£316,288.78
39	oct-16	31	5	£106,901.94	£58,310.15	£308,876.43
40	nov-16	32	6	£1,568,139.35	£855,348.74	£298,403.01
41	dic-16	33	7	£0.00	£506,117.62	£292,175.74
42	ene-17	34	8	£0.00	£506,117.62	£292,287.03
43	mar-17	35	9	£0.00	£506,117.62	£295,614.48
44	mar-17	36	10	£0.00	£506,117.62	£293,820.95
45	abr-17	37	11	£0.00	£506,117.62	£273,354.46
46	may-17	38		£0.00		

Figura 6.8: Ejemplo de dos escenarios proyectados para los presupuestos y tiempos de ejecución de un proyecto. Fuente: Elaboración propia

En la figura 6.8 puede notarse que, en los 2 escenarios los valores de los presupuestos para los periodos anteriores al mes de junio de 2016 son iguales, y dado que hay un mayor número de meses de extensión para el segundo caso, los valores de los presupuestos para los periodos posteriores a la fecha de referencia son menores que en el primer caso.

Pese a ser suficientes los procedimientos anteriores para generar un modelo de simulación completo, es necesaria la construcción de un reporte para sintetizar la información de salida de mayor relevancia de cada uno de los subproyectos del CAP, con el fin de generar estadísticas a nivel general con base en estos resultados.

Un ejemplo de un resumen con los resultados con mayor importancia en el modelo puede incluir los siguientes datos por proyecto: nombre del subproyecto, periodo o fecha de inicio, periodo o fecha de terminación, número de periodos de duración del proyecto, ejecución del proyecto (sí, no), presupuesto total final del subproyecto. El resumen que contiene esta información debe lucir como en la siguiente figura:

Nombre del proyecto	Nombre del subroyecto	Fecha de Inicio	Fecha de Fin	Periodo de Inicio	Periodo de Terminación	Periodos de Duración del Subproyecto	¿Se Inicial el Proyecto? (0=No, 1=Sí)	Presupuesto Total Final del Subproyecto	Presupuesto Total Final del Proyecto
CAP1	CAP1-P1	30/06/2014	30/04/2017	3	37	35	Sí	£25,598,358.97	£65,284,693.06
	CAP1-P2	31/05/2014	30/04/2017	2	37	36	Sí	£3,918,909.75	
	CAP1-P3	01/03/2017	30/04/2018	35	49	15	Sí	£3,187,733.48	
	CAP1-P4	30/04/2014	31/03/2026	1	144	144	Sí	£23,487,797.10	
	CAP1-P5	31/07/2014	30/06/2017	4	39	36	Sí	£9,091,893.76	
CAP2	CAP2-P1	30/04/2015	31/05/2018	13	50	38	Sí	£10,768,346.87	£30,583,234.10
	CAP2-P2	30/06/2015	30/01/2017	15	34	20	Sí	£19,814,887.23	
CAP3	CAP3-P1	31/05/2015	30/05/2018	14	50	37	Sí	£11,829,012.71	£74,048,153.79
	CAP3-P2	30/06/2016	30/06/2019	27	63	37	Sí	£8,310,181.62	
	CAP3-P3	30/09/2015	31/08/2018	18	53	36	Sí	£25,561,779.23	
	CAP3-P4	31/07/2016	31/01/2020	28	70	43	Sí	£28,347,180.23	
	CAP3-P5	31/07/2016	31/05/2019	28	62	35	No	£0.00	
CAP4	CAP4-P1	30/09/2016	30/07/2019	30	64	35	Sí	£5,177,379.40	£69,863,839.66
	CAP4-P2	30/09/2016	30/11/2018	30	56	27	Sí	£45,343,869.95	
	CAP4-P3	31/08/2016	30/04/2018	29	49	21	Sí	£11,480,752.47	
	CAP4-P4	30/11/2016	30/11/2018	32	56	25	No	£0.00	
	CAP4-P5	30/09/2016	30/08/2019	30	65	36	Sí	£5,207,411.25	
	CAP4-P6	31/10/2016	30/10/2019	31	67	37	Sí	£2,654,426.58	
CAP5	CAP5-P1	30/06/2017	31/01/2021	39	82	44	Sí	£58,454,644.22	£58,454,644.22
CAP6	CAP6-P1	30/09/2017	30/04/2021	42	85	44	Sí	£37,308,420.49	£37,308,420.49
CAP7	CAP7-P1	01/03/2018	30/09/2021	47	90	44	Sí	£45,289,233.23	£45,289,233.23
CAP8	CAP8-P1	30/11/2018	30/06/2022	56	99	44	Sí	£45,902,502.67	£45,902,502.67
CAP9	CAP9-P1	31/03/2019	31/10/2022	60	103	44	Sí	£43,731,155.36	£43,731,155.36
CAP10	CAP10-P1	30/09/2019	30/04/2023	66	109	44	Sí	£43,563,765.81	£43,563,765.81
CAP11	CAP11-P1	31/01/2020	31/08/2023	70	113	44	Sí	£44,448,979.99	£44,448,979.99
CAP12	CAP12-P1	31/10/2020	31/05/2024	79	122	44	Sí	£45,363,856.52	£45,363,856.52
CAP13	CAP13-P1	30/04/2021	30/11/2024	85	128	44	Sí	£44,980,722.75	£44,980,722.75
CAP14	CAP14-P1	30/06/2021	31/01/2025	87	130	44	Sí	£46,765,071.31	£46,765,071.31
CAP15	CAP15-P1	30/06/2022	31/01/2026	99	142	44	Sí	£43,322,094.64	£43,322,094.64
CAP16	CAP16-P1	30/11/2022	30/06/2026	104	147	44	Sí	£40,693,495.75	£40,693,495.75
CAP17	CAP17-P1	30/04/2023	31/10/2026	109	151	43	Sí	£45,485,999.37	£45,485,999.37

Figura 6.9: Ejemplo de la estructura del resumen de resultados de los proyectos del CAP. Fuente: Elaboración propia

En la figura 6.9 de hallan 10 columnas, cada una de ellas contiene su respectivo nombre en la parte posterior y los valores de las celdas de la tercera a la décima columna están sujetos a una iteración de la simulación, dado que se encuentra activo el recálculo aleatorio, la última columna consiste en la suma de los presupuestos totales finales de todos los subproyectos que conforman el proyecto correspondiente. Como referencia, la primera celda de la primera columna de esta tabla se encuentra escrita en la celda *B2* de una hoja con el nombre *Resumen* del libro donde se halla el modelo. Y las celdas de la columna *K* son las que se encuentran en color rojo y se han añadido como celdas de salida con el fin de observar el comportamiento de los presupuestos totales de cada proyecto.

Con base en este resumen pueden obtenerse variables y estadísticas que sinteticen la información proporcionada por cada uno de los proyectos. Sin embargo, las 2 variables de mayor importancia en este proyecto, que son las mismas por las que se ha recurrido a utilizar las técnicas de la administración de proyectos y simulación Monte Carlo expuestas en los capítulos anteriores, dichas variables son el presupuesto total del conjunto de proyectos del CAP y su fecha de terminación.

Para calcular el presupuesto total es necesario sumar los montos totales finales de los proyectos o bien de todos los subproyectos, en este caso se realizó con los montos de los subproyectos, utilizando la función *SUMA* cuyos argumentos son los valores de la novena columna de la tabla en la figura 6.9 (Rango *J3:J33*). El periodo y la fecha finales de actividad de los proyectos englobados se calculó por medio de los valores máximos de las columnas sexta y cuarta de la tabla en la figura 6.9 respectivamente. Si los valores del presupuesto total, el periodo final y la fecha final se colocan en las celdas *N3*, *N4* y *N5* respectivamente, y en el mismo orden las fórmulas que contienen son: *=SUMA(J3:J33)*, *=MAX(G3:G33)* y *=MAX(E3:E33)* la presentación de estos resultados luce como en la figura mostrada a continuación, en donde los valores mostrados son de acuerdo con los de la figura 6.9.

	M	N
3	Presupuesto Total	£825,089,862.72
4	Periodo final	151
5	Fecha final	31/10/2026

Figura 6.10: Ejemplo de la presentación del presupuesto total, el periodo y fecha de terminación del conjunto de proyectos del CAP. Fuente: Elaboración propia

Después de haber dado una estructura adecuada al modelo y definir las variables de salida que desean analizarse, solo queda efectuar la simulación del modelo con un número considerable de iteraciones, de modo que se tenga homogeneidad en las distribuciones de las variables de entrada y salida. Debe tomarse en cuenta que se requiere un periodo de tiempo considerable para realizar la simulación, dado el tamaño del modelo y el uso de recursos computacionales que esto implica. En la siguiente sección se mostrarán los resultados de la simulación con 500,000 iteraciones efectuadas.

6.3 Resultados y Conclusiones

Después de efectuar una simulación con un número de 500,000 iteraciones en @Risk, se obtuvieron los siguientes resultados para el presupuesto total, el periodo de terminación y la fecha de terminación de los proyectos a nivel general:

La distribución de probabilidad del Presupuesto Total de todos los proyectos contemplados en el modelo del CAP se muestra en la gráfica de la figura 6.11, al igual que las principales estadísticas referentes a este monto. En el eje horizontal de la gráfica se encuentran los valores del número de millones de libras esterlinas y en el eje vertical la densidad estimada por la muestra generada por la simulación. Como es de esperarse, la distribución asociada a esta variable es acotada por ambos lados, esto se debe a que la mayoría de los proyectos tienen una probabilidad de ejecución del 100%.

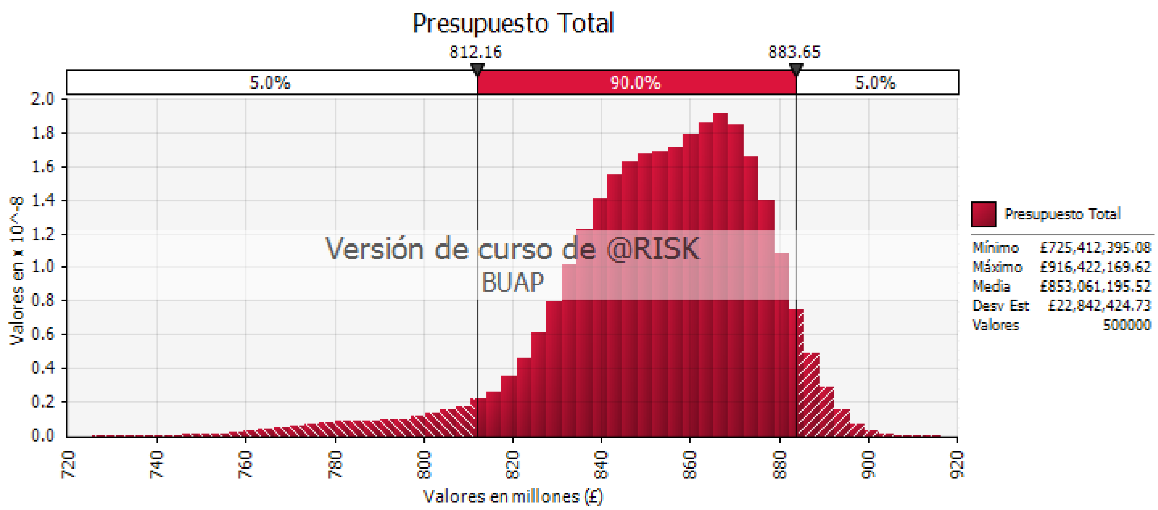

Figura 6.11: Distribución de probabilidad del Presupuesto Total de los proyectos del CAP generada por con una simulación. Fuente: Elaboración propia

En los escenarios generados por la simulación, el presupuesto mínimo requerido para el proyecto es de £725,412,395.08 y el presupuesto máximo requerido es de £916,422,169.62 con una media de £853,061,195.52 y una desviación estándar de £22,842,424.73. Sin embargo, estas estadísticas no son suficientes como apoyo para tomar una decisión certera. En la siguiente ilustración se muestran los valores de la moda de la distribución y sus percentiles:

Moda	£867,644,421.71	50%	£855,827,960.86
1%	£776,825,139.47	55%	£858,739,931.73
5%	£812,156,949.48	60%	£861,507,977.52
10%	£825,931,968.30	65%	£864,237,568.13
15%	£832,330,318.58	70%	£866,861,771.92
20%	£836,790,765.67	75%	£869,470,791.05
25%	£840,462,401.93	80%	£872,206,743.69
30%	£843,769,830.19	85%	£875,213,374.27
35%	£846,868,632.16	90%	£878,745,206.72
40%	£849,893,943.16	95%	£883,645,795.17
45%	£852,880,437.00	99%	£892,298,399.58

Figura 6.12: Resumen estadístico de la simulación. Fuente: Elaboración propia

Los montos presupuestales totales se encuentran alrededor de los 867,644,421.71 de libras esterlinas. El resto de los valores presentados en las columnas de la derecha son los percentiles que corresponden a los porcentajes de las columnas de la izquierda, y estos son útiles para definir el riesgo asumido por los tomadores de decisiones, por ejemplo, se necesita asignar un presupuesto total de £883,645,795.17 para no correr un riesgo mayor al 5% de que el presupuesto asignado sea insuficiente y no se pueda cubrir la totalidad de los proyectos.

A diferencia de la variable asociada con el presupuesto total, la fecha de terminación de los proyectos tiene una distribución de probabilidad discreta, eso se debe a que, pese a tratarse de una variable de tiempo, se refiere a un periodo de un mes específico y no a cualquier punto en el tiempo considerado. Dicha distribución se muestra en la siguiente figura:

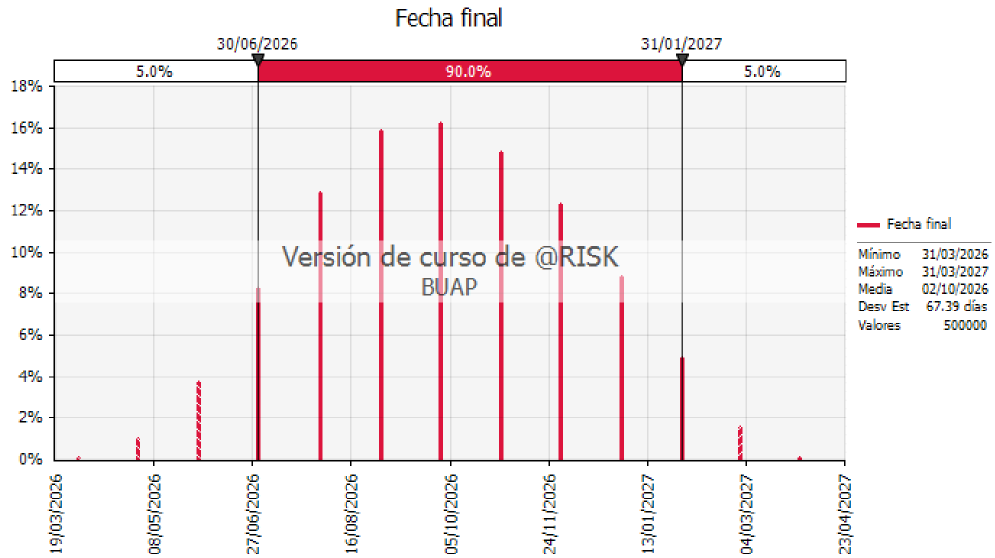

Figura 6.13: Distribución de la fecha de terminación de los proyectos del CAP. Fuente: Elaboración propia

De acuerdo con los escenarios simulados, la fecha más temprana en que se concluirán por completo todos los proyectos incluidos en el modelo es el 31 de marzo del año 2026, sin embargo, esta puede extenderse hasta el 31 de marzo del año 2027. La media de la distribución hace referencia al 2 de octubre del año 2026, sin embargo, en este modelo se están tratando periodos mensuales completos, por esa razón se asumirá el final del mes en que se encuentra la media como la fecha esperada de terminación, es decir, el 31 de octubre del 2026, habiendo una desviación estándar de 67 días. Es importante recordar que a cada fecha se le puede asociar un número de periodo que ayuda a dar una interpretación estadística a las estadísticas relacionadas a esta variable, y como existen proyectos iniciados en el primer periodo contemplado (abril de 2014), y los números de periodo asociados a marzo de 2026 y marzo de 2027 son el 144 y el 156 respectivamente, se puede concluir que se necesita un mínimo de 144 meses (12 años) y un máximo de 156 meses (13 años) para concluir la totalidad de los proyectos.

A pesar de que los tiempos de inicio y conclusión de los proyectos no son manipulables, es importante la obtención de estadísticas acerca de ellos, con el fin de proporcionar a los tomadores de decisiones un panorama de posibilidades que les sea útil para tomar medidas preventivas en caso de surgir desviaciones en sus cronogramas establecidos.

En la siguiente imagen se hallan las principales estadísticas acerca de la fecha de terminación de los proyectos:

Moda	30/09/2026	50%	30/09/2026
1%	31/05/2026	55%	30/09/2026
5%	30/06/2026	60%	31/10/2026
10%	30/06/2026	65%	31/10/2026
15%	31/07/2026	70%	31/10/2026
20%	31/07/2026	75%	30/11/2026
25%	31/07/2026	80%	30/11/2026
30%	31/08/2026	85%	31/12/2026
35%	31/08/2026	90%	31/12/2026
40%	31/08/2026	95%	31/01/2027
45%	30/09/2026	99%	01/03/2027

Figura 6.14: Resumen estadístico de la fecha de terminación de los proyectos del CAP. Fuente: Elaboración propia

En conclusión, los encargados de gestionar y administrar los proyectos tratados en este modelo pueden elegir el riesgo máximo que desean asumir según su experiencia y de esa forma destinar un presupuesto para cubrir la totalidad del proyecto. De la misma forma que con los montos presupuestales, pueden dar continuidad a los tiempos de desarrollo de los proyectos para determinar si se están ejecutando con normalidad o si se presentan desviaciones que deberían ocurrir con poca frecuencia y asumir una conducta de precaución.

Cabe mencionar que el potencial de este modelo no radica en las cifras presentadas anteriormente, sino en el dinamismo que posee para generarlas aun cuando se puedan agregar modificaciones resultantes de las posibles desviaciones que surjan durante el desarrollo de los proyectos. Basta con definir como salida cualquiera de las celdas que contengan información aleatoria en el modelo para observar su comportamiento y obtener estadísticas acerca de la variable asociada.

Anexos

Base de datos de los proyectos del CAP

	CAP1				
	CAP1-P1	CAP1-P2	CAP1-P3	CAP1-P4	CAP1-P5
Periodo Inicial estimado	2	2	35	1	4
Periodo Final Estimado	31	31	49	144	39
Retraso	PERT(0,0,0)	PERT(0,0,0)	PERT(0,0,0)	PERT(0,0,0)	PERT(0,0,0)
Extensión	PERT(0,6,6)	PERT(0,9,12)	PERT(0,0,0)	PERT(0,0,0)	PERT(0,0,0)
Probabilidad de ejecución	1	1	1	1	1
Porcentaje ejercido	PERT(0.75,0.9,1)	PERT(0.9,0.95,1.1)	PERT(0.9,0.95,1.1)	PERT(0.8,0.9,1)	PERT(0.8,0.9,1)
Presupuesto Original	£759,627.29	£137,434.59	£138,040.25	£175,042.21	£111,150.39
	£759,627.29	£137,434.59	£138,040.25	£175,042.21	£111,150.39
	£759,627.29	£137,434.59	£138,040.25	£175,043.21	£111,150.39
	£780,221.87	£138,344.53	£251,694.54	£175,044.21	£212,956.62
	£780,221.87	£138,344.53	£251,694.54	£175,045.21	£212,956.62
	£780,221.87	£138,344.53	£251,694.54	£175,046.21	£212,956.62
	£1,244,012.58	£179,493.41	£266,666.67	£175,047.21	£227,427.34
	£1,244,012.58	£179,493.41	£266,666.67	£175,048.21	£227,427.34
	£1,244,012.58	£179,493.41	£266,666.67	£175,049.21	£227,427.34
	£757,228.83	£170,621.01	£305,333.33	£175,050.21	£248,918.16
	£757,228.83	£170,621.01	£305,333.33	£175,051.21	£248,918.16
	£757,228.83	£170,621.01	£305,333.33	£175,052.21	£248,918.16
	£832,850.77	£180,276.05	£166,666.67	£175,053.21	£198,802.61
	£832,850.77	£180,276.05	£166,666.67	£175,054.21	£198,802.61
	£832,850.77	£180,276.05	£166,666.67	£175,055.21	£198,802.61
	£815,326.08	£197,414.40		£175,056.21	£200,597.57
	£815,326.08	£197,414.40		£175,057.21	£200,597.57
	£815,326.08	£197,414.40		£175,058.21	£200,597.57
	£1,600,641.36	£152,199.39		£175,059.21	£243,229.90
	£1,600,641.36	£152,199.39		£175,060.21	£243,229.90
	£1,600,641.36	£152,199.39		£175,061.21	£243,229.90
	£946,280.67	£94,285.33		£175,062.21	£346,538.87
	£946,280.67	£94,285.33		£175,063.21	£346,538.87
	£946,280.67	£94,285.33		£175,064.21	£346,538.87
	£580,553.33	£14,072.67		£175,065.21	£432,580.81
	£580,553.33	£14,072.67		£175,066.21	£432,580.81
	£580,553.33	£14,072.67		£175,067.21	£432,580.81
	£416,166.67	£29,906.33		£175,068.21	£416,386.39
	£416,166.67	£29,906.33		£175,069.21	£416,386.39
	£416,166.67	£29,906.33		£175,070.21	£416,386.39
				£175,071.21	£389,589.58
				£175,072.21	£389,589.58
				£175,073.21	£389,589.58
				£175,074.21	£262,213.29

				£175,075.21	£262,213.29
				£175,076.21	£262,213.29
				£175,077.21	
				£175,078.21	
				£175,079.21	
				£175,080.21	
				£175,081.21	
				£175,082.21	
				£175,083.21	
				£175,084.21	
				£175,085.21	
				£175,086.21	
				£175,087.21	
				£175,088.21	
				£175,089.21	
				£175,090.21	
				£175,091.21	
				£175,092.21	
				£175,093.21	
				£175,094.21	
				£175,095.21	
				£175,096.21	
				£175,097.21	
				£175,098.21	
				£175,099.21	
				£175,100.21	
				£175,101.21	
				£175,102.21	
				£175,103.21	
				£175,104.21	
				£175,105.21	
				£175,106.21	
				£175,107.21	
				£175,108.21	
				£175,109.21	
				£175,110.21	
				£175,111.21	
				£175,112.21	
				£175,113.21	
				£175,114.21	
				£175,115.21	
				£175,116.21	
				£175,117.21	
				£175,118.21	
				£175,119.21	
				£175,120.21	
				£175,121.21	
				£175,122.21	
				£175,123.21	
				£175,124.21	
				£175,125.21	
				£175,126.21	
				£175,127.21	
				£175,128.21	
				£175,129.21	
				£175,130.21	

				£175,131.21	
				£175,132.21	
				£175,133.21	
				£175,134.21	
				£175,135.21	
				£175,136.21	
				£175,137.21	
				£175,138.21	
				£175,139.21	
				£175,140.21	
				£175,141.21	
				£175,142.21	
				£175,143.21	
				£175,144.21	
				£175,145.21	
				£175,146.21	
				£175,147.21	
				£175,148.21	
				£175,149.21	
				£175,150.21	
				£175,151.21	
				£175,152.21	
				£175,153.21	
				£175,154.21	
				£175,155.21	
				£175,156.21	
				£175,157.21	
				£175,158.21	
				£175,159.21	
				£175,160.21	
				£175,161.21	
				£175,162.21	
				£175,163.21	
				£175,164.21	
				£175,165.21	
				£175,166.21	
				£175,167.21	
				£175,168.21	
				£175,169.21	
				£175,170.21	
				£175,171.21	
				£175,172.21	
				£175,173.21	
				£175,174.21	
				£175,175.21	
				£175,176.21	
				£175,177.21	
				£175,178.21	
				£175,179.21	
				£175,180.21	
				£175,181.21	
				£175,182.21	
				£175,183.21	
				£175,184.21	

	CAP2	
	CAP2-P1	CAP2-P2
Periodo Inicial estimado	13	15
Periodo Final Estimado	48	32
Retraso	PERT(0,0,0)	PERT(0,0,0)
Extensión	PERT(0,0,0)	PERT(0,2,5)
Probabilidad de ejecución	1	1
Porcentaje ejercido	PERT(0.95,0.99,1)	PERT(0.85,0.95,1)
Presupuesto Original	£112,971.16	£4,157,040.96
	£112,971.16	£865,218.19
	£112,971.16	£1,115,060.76
	£169,816.18	£716,552.50
	£169,816.18	£953,839.27
	£169,816.18	£1,330,166.20
	£179,391.61	£1,713,970.66
	£179,391.61	£888,891.67
	£179,391.61	£814,129.28
	£238,243.73	£759,977.32
	£238,243.73	£807,800.38
	£238,243.73	£659,269.72
	£485,861.50	£409,375.40
	£485,861.50	£1,371,027.45
	£485,861.50	£322,982.67
	£381,655.04	£1,788,866.97
	£381,655.04	£106,901.94
	£381,655.04	£1,568,139.35
	£408,108.77	
	£408,108.77	
	£408,108.77	
	£415,507.01	
	£415,507.01	
	£415,507.01	
	£336,256.26	
	£336,256.26	
	£336,256.26	
	£319,736.53	
	£319,736.53	
	£319,736.53	
	£330,158.21	
	£330,158.21	
	£330,158.21	
	£313,092.73	
	£313,092.73	
	£313,092.73	

	CAP3				
	CAP3-P1	CAP3-P2	CAP3-P3	CAP3-P4	CAP3-P5
Periodo Inicial estimado	14	20	18	21	24
Periodo Final Estimado	49	55	53	62	58
Retraso	PERT(0,0,0)	PERT(7,7,12)	PERT(0,0,0)	PERT(6,6,12)	PERT(3,6,12)
Extensión	PERT(0,0,6)	PERT(0,0,6)	PERT(0,0,2)	PERT(0,0,3)	PERT(0,1,7)
Probabilidad de ejecución	1	1	1	1	0.6
Porcentaje ejercido	PERT(0.85,0.95,1)	PERT(0.85,0.95,1)	PERT(0.9,0.95,1)	PERT(0.85,0.95,1)	PERT(0.7,0.9,1)
Presupuesto Original	£126,696.43	£49,150.08	£371,403.18	£251,703.00	£1,016,104.03
	£126,696.43	£49,150.08	£371,403.18	£251,703.00	£508,052.02
	£126,696.43	£49,150.08	£371,403.18	£251,703.00	£957,357.34
	£247,004.18	£143,111.63	£883,478.36	£513,258.33	£957,357.34
	£247,004.18	£143,111.63	£883,478.36	£513,258.33	£957,357.34
	£247,004.18	£143,111.63	£883,478.36	£513,258.33	£1,000,848.34
	£225,002.62	£271,717.31	£860,312.64	£594,997.78	£1,000,848.34
	£225,002.62	£271,717.31	£860,312.64	£594,997.78	£1,000,848.34
	£225,002.62	£271,717.31	£860,312.64	£594,997.78	£1,331,814.01
	£313,648.58	£312,970.16	£845,024.42	£1,209,056.93	£1,331,814.01
	£313,648.58	£312,970.16	£845,024.42	£1,209,056.93	£1,331,814.01
	£313,648.58	£312,970.16	£845,024.42	£1,209,056.93	£846,124.98
	£483,109.02	£376,239.53	£877,760.49	£495,311.30	£846,124.98
	£483,109.02	£376,239.53	£877,760.49	£495,311.30	£846,124.98
	£483,109.02	£376,239.53	£877,760.49	£495,311.30	£1,139,297.66
	£466,721.84	£338,100.19	£768,742.54	£775,415.17	£1,139,297.66
	£466,721.84	£338,100.19	£768,742.54	£775,415.17	£1,139,297.66
	£466,721.84	£338,100.19	£768,742.54	£775,415.17	£806,453.91
	£543,222.26	£350,977.16	£981,221.78	£1,377,204.78	£806,453.91
	£543,222.26	£350,977.16	£981,221.78	£1,377,204.78	£806,453.91
	£543,222.26	£350,977.16	£981,221.78	£1,377,204.78	£925,252.23
	£470,815.88	£349,650.66	£971,374.21	£872,338.51	£925,252.23
	£470,815.88	£349,650.66	£971,374.21	£872,338.51	£925,252.23
	£470,815.88	£349,650.66	£971,374.21	£872,338.51	£772,219.63
	£396,834.38	£261,483.19	£884,492.26	£860,697.95	£772,219.63
	£396,834.38	£261,483.19	£884,492.26	£860,697.95	£772,219.63
	£396,834.38	£261,483.19	£884,492.26	£860,697.95	£643,717.97
	£399,310.67	£198,666.32	£484,748.10	£379,368.30	£643,717.97
	£399,310.67	£198,666.32	£484,748.10	£379,368.30	£643,717.97
	£399,310.67	£198,666.32	£484,748.10	£379,368.30	£595,934.43
	£323,795.54	£138,629.02	£412,421.64	£642,146.98	£595,934.43
	£323,795.54	£138,629.02	£412,421.64	£642,146.98	£595,934.43
	£323,795.54	£138,629.02	£412,421.64	£642,146.98	£464,117.29
	£269,143.60	£77,369.99	£293,822.64	£641,772.51	£464,117.29
	£269,143.60	£77,369.99	£293,822.64	£641,772.51	£464,117.29
	£269,143.60	£77,369.99	£293,822.64	£641,772.51	
				£614,407.65	
				£614,407.65	
				£614,407.65	
				£485,490.81	
				£485,490.81	
				£485,490.81	

	CAP4					
	CAP4-P1	CAP4-P2	CAP4-P3	CAP4-P4	CAP4-P5	CAP4-P6
Periodo Inicial estimado	25	25	25	24	26	25
Periodo Final Estimado	57	48	45	46	61	59
Retraso	PERT(2,6,12)	PERT(2,6,12)	PERT(2,6,12)	PERT(3,6,12)	PERT(1,6,12)	PERT(2,6,12)
Extensión	PERT(0,3,9)	PERT(0,0,6)	PERT(0,0,5)	PERT(0,3,7)	PERT(0,0,4)	PERT(0,1,7)
Probabilidad de ejecución	1	1	0.85	0.85	0.9	0.9
Porcentaje ejercido	PERT(0.8,0.95,1)	PERT(0.85,0.95,1)	PERT(0.8,0.9,1)	PERT(0.8,0.9,1)	PERT(0.8,0.9,1)	PERT(0.8,0.9,1)
Presupuesto Original	£409,182.32	£544,448.25	£174,962.79	£5,736,000.00	£90,035.97	£95,263.83
	£153,153.32	£208,448.25	£174,962.79	£1,881,100.00	£90,035.97	£47,631.92
	£153,153.32	£208,448.25	£174,962.79	£755,307.80	£90,035.97	£66,684.68
	£201,633.83	£1,154,923.95	£481,147.66	£755,307.80	£126,050.35	£66,684.68
	£201,633.83	£1,154,923.95	£481,147.66	£755,307.80	£126,050.35	£66,684.68
	£201,633.83	£1,154,923.95	£481,147.66	£843,960.64	£126,050.35	£85,737.45
	£225,380.78	£2,491,378.55	£656,110.45	£843,960.64	£162,064.74	£85,737.45
	£225,380.78	£2,491,378.55	£656,110.45	£843,960.64	£162,064.74	£85,737.45
	£225,380.78	£2,491,378.55	£656,110.45	£926,781.58	£162,064.74	£76,211.07
	£198,262.79	£3,086,338.14	£831,073.24	£926,781.58	£144,057.55	£76,211.07
	£198,262.79	£3,086,338.14	£831,073.24	£926,781.58	£144,057.55	£76,211.07
	£198,262.79	£3,086,338.14	£831,073.24	£801,034.31	£144,057.55	£85,737.45
	£181,923.93	£2,621,020.59	£1,268,480.20	£801,034.31	£162,064.74	£85,737.45
	£181,923.93	£2,621,020.59	£1,268,480.20	£801,034.31	£162,064.74	£85,737.45
	£181,923.93	£2,621,020.59	£1,268,480.20	£320,210.40	£162,064.74	£95,263.83
	£182,754.21	£2,200,361.51	£831,073.24	£320,210.40	£180,071.93	£95,263.83
	£182,754.21	£2,200,361.51	£831,073.24	£320,210.40	£180,071.93	£95,263.83
	£182,754.21	£2,200,361.51	£831,073.24	£86,875.74	£180,071.93	£123,842.98
	£185,275.86	£2,072,959.83	£174,962.79	£86,875.74	£234,093.51	£123,842.98
	£185,275.86	£2,072,959.83	£174,962.79	£86,875.74	£234,093.51	£123,842.98
	£185,275.86	£2,072,959.83	£174,962.79	£68,578.24	£234,093.51	£104,790.22
	£181,861.37	£1,544,388.43		£68,578.24	£198,079.13	£104,790.22
	£181,861.37	£1,544,388.43		£68,578.24	£198,079.13	£104,790.22
	£181,861.37	£1,544,388.43			£198,079.13	£95,263.83
	£165,574.85				£180,071.93	£95,263.83
	£165,574.85				£180,071.93	£95,263.83
	£165,574.85				£180,071.93	£95,263.83
	£131,535.11				£180,071.93	£95,263.83
	£131,535.11				£180,071.93	£95,263.83
	£131,535.11				£180,071.93	£57,158.30
	£95,475.64				£108,043.16	£57,158.30
	£95,475.64				£108,043.16	£57,158.30
	£95,475.64				£108,043.16	£19,052.77
					£36,014.39	£19,052.77
					£36,014.39	£19,052.77
					£36,014.39	

	CAP5	CAP6	CAP7	CAP8	CAP9
	CAP5-P1	CAP6-P1	CAP7-P1	CAP8-P1	CAP9-P1
Periodo Inicial estimado	30	36	42	48	54
Periodo Final Estimado	73	79	85	91	97
Retraso	PERT(0,6,12)	PERT(0,6,12)	PERT(0,6,12)	PERT(0,6,12)	PERT(0,6,12)
Extensión	PERT(0,0,0)	PERT(0,0,0)	PERT(0,0,0)	PERT(0,0,0)	PERT(0,0,0)
Probabilidad de ejecución	0.95	1	1	1	1
Porcentaje ejercido	PERT(0.8,0.9,1)	PERT(0.5,0.6,1)	PERT(0.75,0.9,1)	PERT(0.75,0.9,1)	PERT(0.75,0.9,1)
Presupuesto Original	£179,482.00	£166,661.86	£128,201.43	£128,201.43	£128,201.43
	£430,934.39	£400,153.36	£307,810.28	£307,810.28	£307,810.28
	£673,469.09	£625,364.16	£481,049.35	£481,049.35	£481,049.35
	£905,767.55	£841,069.87	£646,976.82	£646,976.82	£646,976.82
	£1,126,622.94	£1,046,149.87	£804,730.67	£804,730.67	£804,730.67
	£1,334,940.15	£1,239,587.28	£953,528.68	£953,528.68	£953,528.68
	£1,529,735.80	£1,420,468.96	£1,092,668.43	£1,092,668.43	£1,092,668.43
	£1,710,138.22	£1,587,985.49	£1,221,527.30	£1,221,527.30	£1,221,527.30
	£1,875,387.46	£1,741,431.21	£1,339,562.47	£1,339,562.47	£1,339,562.47
	£2,024,835.29	£1,880,204.20	£1,446,310.92	£1,446,310.92	£1,446,310.92
	£2,157,945.20	£2,003,806.26	£1,541,389.43	£1,541,389.43	£1,541,389.43
	£2,274,292.41	£2,111,842.95	£1,624,494.58	£1,624,494.58	£1,624,494.58
	£2,373,563.85	£2,204,023.58	£1,695,402.75	£1,695,402.75	£1,695,402.75
	£2,455,558.17	£2,280,161.16	£1,753,970.12	£1,753,970.12	£1,753,970.12
	£2,520,185.74	£2,340,172.47	£1,800,132.67	£1,800,132.67	£1,800,132.67
	£2,567,468.65	£2,384,078.03	£1,833,906.18	£1,833,906.18	£1,833,906.18
	£2,597,540.72	£2,412,002.10	£1,855,386.23	£1,855,386.23	£1,855,386.23
	£2,610,647.48	£2,424,172.66	£1,864,748.20	£1,864,748.20	£1,864,748.20
	£2,607,146.18	£2,420,921.45	£1,862,247.27	£1,862,247.27	£1,862,247.27
	£2,587,505.79	£2,402,683.95	£1,848,218.42	£1,848,218.42	£1,848,218.42
	£2,552,307.00	£2,369,999.36	£1,823,076.43	£1,823,076.43	£1,823,076.43
	£2,502,242.23	£2,323,510.64	£1,787,315.88	£1,787,315.88	£1,787,315.88
	£2,438,115.61	£2,263,964.50	£1,741,511.15	£1,741,511.15	£1,741,511.15
	£2,360,842.99	£2,192,211.35	£1,686,316.42	£1,686,316.42	£1,686,316.42
	£2,271,451.94	£2,109,205.37	£1,622,465.67	£1,622,465.67	£1,622,465.67
	£2,171,081.75	£2,016,004.48	£1,550,772.68	£1,550,772.68	£1,550,772.68
	£2,060,983.44	£1,913,770.34	£1,472,131.03	£1,472,131.03	£1,472,131.03
	£1,942,519.74	£1,803,768.33	£1,387,514.10	£1,387,514.10	£1,387,514.10
	£1,817,165.10	£1,687,367.59	£1,297,975.07	£1,297,975.07	£1,297,975.07
	£1,686,505.69	£1,566,041.00	£1,204,646.92	£1,204,646.92	£1,204,646.92
	£1,552,239.40	£1,441,365.16	£1,108,742.43	£1,108,742.43	£1,108,742.43
	£1,416,175.85	£1,315,020.43	£1,011,554.18	£1,011,554.18	£1,011,554.18
	£1,280,236.37	£1,188,790.92	£914,454.55	£914,454.55	£914,454.55
	£1,146,454.01	£1,064,564.44	£818,895.72	£818,895.72	£818,895.72
	£1,016,973.54	£944,332.57	£726,409.67	£726,409.67	£726,409.67
	£894,051.45	£830,190.63	£638,608.18	£638,608.18	£638,608.18
	£780,055.96	£724,337.68	£557,182.83	£557,182.83	£557,182.83
	£677,467.00	£629,076.50	£483,905.00	£483,905.00	£483,905.00
	£588,876.22	£546,813.63	£420,625.87	£420,625.87	£420,625.87
	£516,986.99	£480,059.35	£369,276.42	£369,276.42	£369,276.42
	£464,614.40	£431,427.66	£331,867.43	£331,867.43	£331,867.43
	£434,685.27	£403,636.32	£310,489.48	£310,489.48	£310,489.48
	£430,238.13	£399,506.84	£307,312.95	£307,312.95	£307,312.95
	£454,423.23	£421,964.43	£324,588.02	£324,588.02	£324,588.02

	CAP10	CAP11	CAP12	CAP13	CAP14
	CAP10-P1	CAP11-P1	CAP12-P1	CAP13-P1	CAP14-P1
Periodo Inicial estimado	60	66	72	78	84
Periodo Final Estimado	103	109	115	121	127
Retraso	PERT(0,6,12)	PERT(0,6,12)	PERT(0,6,12)	PERT(0,6,12)	PERT(0,6,12)
Extensión	PERT(0,0,0)	PERT(0,0,0)	PERT(0,0,0)	PERT(0,0,0)	PERT(0,0,0)
Probabilidad de ejecución	1	1	1	1	1
Porcentaje ejercido	PERT(0.75,0.9,1)	PERT(0.75,0.9,1)	PERT(0.75,0.9,1)	PERT(0.75,0.9,1)	PERT(0.75,0.9,1)
Presupuesto Original	£128,201.43	£128,201.43	£128,201.43	£128,201.43	£128,201.43
	£307,810.28	£307,810.28	£307,810.28	£307,810.28	£307,810.28
	£481,049.35	£481,049.35	£481,049.35	£481,049.35	£481,049.35
	£646,976.82	£646,976.82	£646,976.82	£646,976.82	£646,976.82
	£804,730.67	£804,730.67	£804,730.67	£804,730.67	£804,730.67
	£953,528.68	£953,528.68	£953,528.68	£953,528.68	£953,528.68
	£1,092,668.43	£1,092,668.43	£1,092,668.43	£1,092,668.43	£1,092,668.43
	£1,221,527.30	£1,221,527.30	£1,221,527.30	£1,221,527.30	£1,221,527.30
	£1,339,562.47	£1,339,562.47	£1,339,562.47	£1,339,562.47	£1,339,562.47
	£1,446,310.92	£1,446,310.92	£1,446,310.92	£1,446,310.92	£1,446,310.92
	£1,541,389.43	£1,541,389.43	£1,541,389.43	£1,541,389.43	£1,541,389.43
	£1,624,494.58	£1,624,494.58	£1,624,494.58	£1,624,494.58	£1,624,494.58
	£1,695,402.75	£1,695,402.75	£1,695,402.75	£1,695,402.75	£1,695,402.75
	£1,753,970.12	£1,753,970.12	£1,753,970.12	£1,753,970.12	£1,753,970.12
	£1,800,132.67	£1,800,132.67	£1,800,132.67	£1,800,132.67	£1,800,132.67
	£1,833,906.18	£1,833,906.18	£1,833,906.18	£1,833,906.18	£1,833,906.18
	£1,855,386.23	£1,855,386.23	£1,855,386.23	£1,855,386.23	£1,855,386.23
	£1,864,748.20	£1,864,748.20	£1,864,748.20	£1,864,748.20	£1,864,748.20
	£1,862,247.27	£1,862,247.27	£1,862,247.27	£1,862,247.27	£1,862,247.27
	£1,848,218.42	£1,848,218.42	£1,848,218.42	£1,848,218.42	£1,848,218.42
	£1,823,076.43	£1,823,076.43	£1,823,076.43	£1,823,076.43	£1,823,076.43
	£1,787,315.88	£1,787,315.88	£1,787,315.88	£1,787,315.88	£1,787,315.88
	£1,741,511.15	£1,741,511.15	£1,741,511.15	£1,741,511.15	£1,741,511.15
	£1,686,316.42	£1,686,316.42	£1,686,316.42	£1,686,316.42	£1,686,316.42
	£1,622,465.67	£1,622,465.67	£1,622,465.67	£1,622,465.67	£1,622,465.67
	£1,550,772.68	£1,550,772.68	£1,550,772.68	£1,550,772.68	£1,550,772.68
	£1,472,131.03	£1,472,131.03	£1,472,131.03	£1,472,131.03	£1,472,131.03
	£1,387,514.10	£1,387,514.10	£1,387,514.10	£1,387,514.10	£1,387,514.10
	£1,297,975.07	£1,297,975.07	£1,297,975.07	£1,297,975.07	£1,297,975.07
	£1,204,646.92	£1,204,646.92	£1,204,646.92	£1,204,646.92	£1,204,646.92
	£1,108,742.43	£1,108,742.43	£1,108,742.43	£1,108,742.43	£1,108,742.43
	£1,011,554.18	£1,011,554.18	£1,011,554.18	£1,011,554.18	£1,011,554.18
	£914,454.55	£914,454.55	£914,454.55	£914,454.55	£914,454.55
	£818,895.72	£818,895.72	£818,895.72	£818,895.72	£818,895.72
	£726,409.67	£726,409.67	£726,409.67	£726,409.67	£726,409.67
	£638,608.18	£638,608.18	£638,608.18	£638,608.18	£638,608.18
	£557,182.83	£557,182.83	£557,182.83	£557,182.83	£557,182.83
	£483,905.00	£483,905.00	£483,905.00	£483,905.00	£483,905.00
	£420,625.87	£420,625.87	£420,625.87	£420,625.87	£420,625.87
	£369,276.42	£369,276.42	£369,276.42	£369,276.42	£369,276.42
	£331,867.43	£331,867.43	£331,867.43	£331,867.43	£331,867.43
	£310,489.48	£310,489.48	£310,489.48	£310,489.48	£310,489.48
	£307,312.95	£307,312.95	£307,312.95	£307,312.95	£307,312.95
	£324,588.02	£324,588.02	£324,588.02	£324,588.02	£324,588.02

	CAP15	CAP16	CAP17
	CAP15-P1	CAP16-P1	CAP17-P1
Periodo Inicial estimado	90	96	102
Periodo Final Estimado	133	139	144
Retraso	PERT(0,6,12)	PERT(0,6,12)	PERT(0,6,12)
Extensión	PERT(0,0,0)	PERT(0,0,0)	PERT(0,0,0)
Probabilidad de ejecución	1	1	1
Porcentaje ejercido	PERT(0.75,0.9,1)	PERT(0.75,0.9,1)	PERT(0.75,0.9,1)
Presupuesto Original	£128,201.43	£128,201.43	£128,201.43
	£307,810.28	£307,810.28	£307,810.28
	£481,049.35	£481,049.35	£481,049.35
	£646,976.82	£646,976.82	£646,976.82
	£804,730.67	£804,730.67	£804,730.67
	£953,528.68	£953,528.68	£953,528.68
	£1,092,668.43	£1,092,668.43	£1,092,668.43
	£1,221,527.30	£1,221,527.30	£1,221,527.30
	£1,339,562.47	£1,339,562.47	£1,339,562.47
	£1,446,310.92	£1,446,310.92	£1,446,310.92
	£1,541,389.43	£1,541,389.43	£1,541,389.43
	£1,624,494.58	£1,624,494.58	£1,624,494.58
	£1,695,402.75	£1,695,402.75	£1,695,402.75
	£1,753,970.12	£1,753,970.12	£1,753,970.12
	£1,800,132.67	£1,800,132.67	£1,800,132.67
	£1,833,906.18	£1,833,906.18	£1,833,906.18
	£1,855,386.23	£1,855,386.23	£1,855,386.23
	£1,864,748.20	£1,864,748.20	£1,864,748.20
	£1,862,247.27	£1,862,247.27	£1,862,247.27
	£1,848,218.42	£1,848,218.42	£1,848,218.42
	£1,823,076.43	£1,823,076.43	£1,823,076.43
	£1,787,315.88	£1,787,315.88	£1,787,315.88
	£1,741,511.15	£1,741,511.15	£1,741,511.15
	£1,686,316.42	£1,686,316.42	£1,686,316.42
	£1,622,465.67	£1,622,465.67	£1,622,465.67
	£1,550,772.68	£1,550,772.68	£1,550,772.68
	£1,472,131.03	£1,472,131.03	£1,472,131.03
	£1,387,514.10	£1,387,514.10	£1,387,514.10
	£1,297,975.07	£1,297,975.07	£1,297,975.07
	£1,204,646.92	£1,204,646.92	£1,204,646.92
	£1,108,742.43	£1,108,742.43	£1,108,742.43
	£1,011,554.18	£1,011,554.18	£1,011,554.18
	£914,454.55	£914,454.55	£914,454.55
	£818,895.72	£818,895.72	£818,895.72
	£726,409.67	£726,409.67	£726,409.67
	£638,608.18	£638,608.18	£638,608.18
	£557,182.83	£557,182.83	£557,182.83
	£483,905.00	£483,905.00	£483,905.00
	£420,625.87	£420,625.87	£420,625.87
	£369,276.42	£369,276.42	£369,276.42
	£331,867.43	£331,867.43	£331,867.43
	£310,489.48	£310,489.48	£310,489.48
	£307,312.95	£307,312.95	£307,312.95
	£324,588.02	£324,588.02	

Código para generar una muestra con distribución Bernoulli(p) en VBA

```
Sub simulaBernoulli()
 Cells.Clear
 p = Val(InputBox("Introduzca el valor de p", "Probabilidad de éxito"))
 n = Val(InputBox("Introduce el número de valores a simular", "Tamaño de muestra"))

 For i = 1 To n
  If Rnd() <= p Then
   x = 1
  Else: x = 0
  End If

  Cells(i + 2, 3) = x
 Next i

End Sub
```

Código para generar una muestra con distribución Uniforme(a,b) en VBA

```
Sub simulaUniforme
  Cells.Clear
  a = Val(InputBox("Introduzca el valor del parámetro a", "Mínimo"))
  b = Val(InputBox("Introduzca el valor del parámetro b", "Máximo"))
  n = Val(InputBox("Introduce el número de valores a simular", "Tamaño de muestra"))

  For i = 1 To n
    Cells(i + 2, 3) = (b - a) * Rnd() + a
  Next i

End Sub
```

Código para generar una muestra con distribución Triangular(a,m,b) en VBA

```
Sub simulaTriangular()
  Cells.Clear
  a = Val(InputBox("Introduce el valor mínimo", "Mínimo"))
  b = Val(InputBox("Introduce el valor máximo", "Máximo"))
  m = Val(InputBox("Introduce el valor más probable", "Moda"))
  n = Val(InputBox("Introduce el número de valores a simular", "Tamaño de muestra"))
  k = 0
  señal = False
  maxf = 2 / (b - a)
  Do
    x = (b - a) * Rnd() + a
    uh = maxf * Rnd()
    If (x <= m And uh <= 2 * (x - a) / ((m - a) * (b - a))) Or (x > m And uh <= 2 * (b - x) / ((b -
m) * (b - a))) Then
      k = k + 1
      Cells(k + 2, 3) = x
    End If

    If k = n Then
      señal = True
    End If

  Loop While señal = False

End Sub
```

Código para generar una muestra con distribución PERT(a,m,b) en VBA

```
Sub simulaPert()
 Cells.Clear
 a = Val(InputBox("Introduzca el valor mínimo", "Mínimo"))
 b = Val(InputBox("Introduzca el valor máximo", "Máximo"))
 m = Val(InputBox("Introduzca el valor más probable", "Moda"))
 n = Val(InputBox("Introduzca el número de valores a simular", "Tamaño de muestra"))
 k = 0
 señal = False
 v1 = 0
 v2 = v1 + 1 / 10000
 w1 = 0
 w2 = w2 + 1 / 10000
 p1 = (4 * m + b - 5 * a) / (b - a)
 p2 = (5 * b - a - 4 * m) / (b - a)
 s1 = 0
 s2 = 0

 For i = 1 To 1000000
  h11 = Exp(-v1) * (v1 ^ (p1 - 1))
  h12 = Exp(-v2) * (v2 ^ (p1 - 1))
  ms1 = (h11 + h12) * (1 / 1000) / 2
  v1 = v1 + 1 / 10000
  v2 = v2 + 1 / 10000
  s1 = s1 + ms1
 Next i

 For j = 1 To 1000000
```

```
    h21 = Exp(-w1) * (w1 ^ (p2 - 1))
    h22 = Exp(-w2) * (w2 ^ (p2 - 1))
    ms2 = (h21 + h22) * (1 / 1000) / 2
    w1 = w1 + 1 / 10000
    w2 = w2 + 1 / 10000
    s2 = s2 + ms2
  Next j

  maxf = (120 * ((m - a) ^ (4 * (m - a) / (b - a))) * ((b - m) ^ (4 * (b - m) / (b - a)))) / (((b - a) ^
5) * s1 * s2)

  Do
    uh = maxf * Rnd()
    x = (b - a) * Rnd() + a
    If uh <= (120 * ((x - a) ^ (4 * (m - a) / (b - a))) * ((b - x) ^ (4 * (b - m) / (b - a)))) / (((b - a) ^
5) * s1 * s2) Then
      k = k + 1
      Cells(k + 2, 3) = x
    End If

    If k = n Then
      señal = True
    End If
  Loop While señal = False


End Sub
```

Bibliografía

Alexander, M., & Kusleika, D. (2016). Excel 2016 Formulas. Indianapolis, EEUU: John Wiley & Sons, Inc.

Esteban González, M. V. (08 de 2009). *Econometría básica Aplicada con Gretl.* Recuperado el 16 de 05 de 2017, de Universidad del País Vasco: https://addi.ehu.es/handle/10810/12496

Ezequiel, U. (Septiembre de 2013). *Universidad de Valencia.* Recuperado el 17 de Marzo de 2017, de http://www.uv.es/=uriel/3%20Regresion%20lineal%20multiple%20estimacion%20y%20propiedades.pdf

Mason, S., Hill, R., Mönch, L., Rose, O., & Jefferson, T. (2008). *Proceedings of the 2008 Winter Simulation Conference.* Fowler eds.

Microsoft. (s.f.). *Función SI*. Recuperado el 21 de Agosto de 2017, de https://support.office.com/es-es/article/Funci%C3%B3n-SI-69AED7C9-4E8A-4755-A9BC-AA8BBFF73BE2?NS=EXCEL&Version=16&SysLcid=3082&UiLcid=3082&AppVer=ZXL160&HelpId=xlmain11.chm60049&ui=es-ES&rs=es-ES&ad=ES

Microsoft. (s.f.). *Función MAX*. Recuperado el 21 de Agosto de 2017, de https://support.office.com/es-es/article/MAX-funci%C3%B3n-MAX-e0012414-9ac8-4b34-9a47-73e662c08098?NS=EXCEL&Version=16&SysLcid=3082&UiLcid=3082&AppVer=ZXL160&HelpId=xlmain11.chm60055&ui=es-ES&rs=es-ES&ad=ES

Microsoft. (s.f.). *Función Y*. Recuperado el 21 de Agosto de 2017, de https://support.office.com/es-es/article/Y-funci%C3%B3n-Y-5f19b2e8-e1df-4408-897a-ce285a19e9d9?NS=EXCEL&Version=16&SysLcid=3082&UiLcid=3082&AppVer=ZXL160&HelpId=xlmain11.chm60084&ui=es-ES&rs=es-ES&ad=ES

Microsoft. (s.f.). *Función SUMA*. Recuperado el 21 de Agosto de 2017, de https://support.office.com/es-es/article/Funci%C3%B3n-SUMA-043e1c7d-7726-4e80-8f32-07b23e057f89?NS=EXCEL&Version=16&SysLcid=3082&UiLcid=3082&AppVer=ZXL160&HelpId=xlmain11.chm60052&ui=es-ES&rs=es-ES&ad=ES

Microsoft. (s.f.). *Función REDONDEAR*. Recuperado el 21 de Agosto de 2017, de https://support.office.com/es-es/article/REDONDEAR-funci%C3%B3n-REDONDEAR-c018c5d8-40fb-4053-90b1-b3e7f61a213c?NS=EXCEL&Version=16&SysLcid=3082&UiLcid=3082&AppVer=ZXL160&HelpId=xlmain11.chm60075&ui=es-ES&rs=es-ES&ad=ES

Microsoft. (s.f.). *Función DESREF*. Recuperado el 21 de Agosto de 2017, de https://support.office.com/es-es/article/DESREF-funci%C3%B3n-DESREF-c8de19ae-dd79-4b9b-a14e-b4d906d11b66?NS=EXCEL&Version=16&SysLcid=3082&UiLcid=3082&AppVer=ZXL160&HelpId=xlmain11.chm60126&ui=es-ES&rs=es-ES&ad=ES

Mukhopdhyay, N. (2000). Probability and Statistical Inference. Nueva York, Nueva York, Estados Unidos: Marcel Dekker, Inc.

Newton, P. (2015). Managing Project Risk. Free Management Ebooks.

Palisade Decision Tools @Risk Manual. (s.f.).

Raychaudhuri, S. (s.f.). *Introduction to Monte Carlo simulation.* Broomfield, EEUU: Oracle Crystal Ball Global Business Unit.

Ross, S. M. (2010). A first course in probability. Upper Saddle, Nueva Jersey, Estados Unidos: Pearson.

Rubinstein, R., & Kroese, D. (2008). Simulation and the Monte Carlo Method. Hoboken, EEUU: John Wiley & Sons, Inc.

The beta-PERT distribution. (s.f.). *RiskAMP.* Obtenido de RiskAMP: 23 de Abril, 2017 https://www.riskamp.com/beta-pert

Torres Hernández, Z., & Torres Martínez, H. (2014). Administración de proyectos. México, D.F.: Grupo Editorial Patria.

Uniform distribution. (s.f.). *Epix Analytics.* Obtenido de Epix Analytics: 04 de Mayo, 2017 http://www.epixanalytics.com/modelassist/AtRisk/Model_Assist.htm#Modeling_expert_opinions/Uniform_distribution.htm

Wackerly, D. D., Mendenhall, W., & Schaffer, R. (2008). Estadística matemática con aplicaciones. Ciudad de México, México: Cengage Learning Editores, S.A.