



APLICACIÓN INFORMÁTICA PARA PRUEBAS ESTADÍSTICAS DE BONDAD DE AJUSTE

Marisol Barrón Bastida
Eric Moreno Quintero

**Publicación Técnica No. 577
Sanfandila, Qro. 2020**

SECRETARÍA DE COMUNICACIONES Y TRANSPORTES
INSTITUTO MEXICANO DEL TRANSPORTE

**Aplicación informática para pruebas estadísticas de
bondad de ajuste**

Publicación Técnica No. 577
Sanfandila, Qro. 2020

Esta aplicación fue realizada en la Coordinación de Transporte Integrado y Logística del Instituto Mexicano del Transporte (IMT), por los investigadores: Mtra. Marisol Barrón Bastida y el Dr. Eric Moreno Quintero.

Este este documento es el informe final del proyecto de investigación interna TI 21/18” Aplicación informática para pruebas estadísticas de bondad de ajuste” realizado en la Coordinación de Transporte Integrado y Logística.

Se agradece la revisión y aportaciones del Dr. Carlos Martner Peyrelongue, coordinador de Transporte Integrado y Logística del IMT, cuyas observaciones mejoraron la calidad de este documento.

Se reconoce también el apoyo brindado por el Laboratorio Nacional de Sistemas de Transporte y Logística (SiT-LOG Lab), cuyas capacidades tecnológicas facilitaron el desarrollo de la aplicación, así como su difusión en el sitio web del SiT-LOG Lab.

Contenido

Índice de figuras	v
Índice de tablas	vii
Sinopsis.....	ix
Abstract	xi
Resumen ejecutivo	xiii
1 Antecedentes.....	1
2 Desarrollo de la aplicación informática	3
2.1 El modelo iterativo incremental	3
2.2 Fases estructurales.....	4
2.2.1 Fase I. Comunicación	4
2.2.1.1 Objetivo	4
2.2.1.2 Alcances.....	4
2.2.1.3 Requerimientos	4
2.2.2 Fase II. Planeación	8
2.2.3 Fase III. Modelado	11
2.2.3.1 Casos de uso de la aplicación.....	11
2.2.3.2 Modelo de secuencia de los datos	14
2.2.3.3 Diagramas de clases.....	15
2.2.4 Fase IV. Construcción	21
2.2.4.1 Pruebas de contenido	21
2.2.4.2 Pruebas de interfaz de usuario	22
2.2.4.3 Pruebas en el nivel de componente	22

2.2.4.4 Prueba de configuración	23
2.2.4.5 Pruebas de seguridad	24
2.2.4.6 Pruebas de rendimiento	24
2.2.5 Fase V. Despliegue	25
3 Distribuciones de probabilidad en transporte e ingeniería	31
3.1 Distribuciones discretas	31
3.2 Distribuciones continuas	35
4 Técnicas de Bondad de Ajuste	41
4.1 La prueba Ji-Cuadrada	43
4.2 La prueba de Kolmogorov-Smirnov	45
4.3 La prueba de Anderson-Darling	47
5 Ejemplos de uso de la aplicación con datos reales	49
6 Resultados y conclusiones	55
Bibliografía	57

Índice de figuras

Figura 2.2 Diagrama de caso de uso: pruebas estadísticas de bondad de ajuste.	13
Figura 2.3 Diagrama de caso de uso: modelado de la distribución binomial.....	14
Figura 2.4 Secuencia de procesamiento de los datos.....	15
Figura 2.5 Diagrama de clases para las distribuciones discretas.....	16
Figura 2.6 Diagrama de clases distribución binomial	17
Figura 2.7 Diagrama de clases distribución Geométrica (pruebas y fallas).....	17
Figura 2.8 Diagrama de clases distribución Poisson.....	18
Figura 2.9 Diagrama de clases distribución Uniforme	18
Figura 2.10 Diagrama de clases distribución Normal	19
Figura 2.11 Diagrama de clases distribución Exponencial	19
Figura 2.12 Diagrama de clases distribución Lognormal.....	20
Figura 2.13 Diagrama de clases distribución Weibull.....	20
Figura 2.14 Diagrama de clases distribución Gamma.....	20
Figura 2.15 Página principal de ABAL.....	26
Figura 2.16 Página principal de ABAL.....	27
Figura 2.17 Sección de ingreso de la muestra de datos.	27
Figura 2.18 Sección de ingreso de la muestra de datos.	28
Figura 2.19 Sección de resultados	29
Figura 2.20 Descarga de documento pdf.	30
Figura 2.21 Vista del documento Binomial.pdf	30
Figura 3.1 Ejemplos de distribución Binomial (elaboración propia)	33
Figura 3.2 Ejemplos de distribución Geométrica	33

Figura 3.3 Ejemplos de distribución Pascal.....	34
Figura 3.4 Ejemplos de distribución Poisson.....	34
Figura 3.5 Ejemplos de distribución Logarítmica.....	35
Figura 3.6 Ejemplos de distribución Normal.....	37
Figura 3.7 Ejemplos de distribución Exponencial.....	37
Figura 3.8 Ejemplos de distribución Gamma.....	38
Figura 3.9 Ejemplos de distribución Lognormal.....	38
Figura 3.10 Ejemplos de distribución Weibull.....	39
Figura 3.11 Ejemplos de distribución Logística.....	39
Figura 3.12 Ejemplos de distribución Gumbel.....	40
Figura 5.1 Aforo vehicular. Puente Internacional Nuevo Laredo (2015). (elaboración propia con ABAL).....	49
Figura 5.2 Minutos entre llegadas de Aeroméxico a CDMX (2013). (elaboración propia con ABAL).....	51
Figura 5.3 Peso del eje motriz. Camiones T3S2, Estación 246/2010 (elaboración propia con ABAL).....	53

Índice de tablas

Tabla 2.1 Requerimiento 1. Diseño de la interfaz de la página principal.....	5
Tabla 2.2 Requerimiento 2 - Diseño de las interfaces para el ingreso de datos en la opción de distribuciones discretas.....	5
Tabla 2.3 Requerimiento 3 - Diseño de las interfaces para el ingreso de datos en la opción de distribuciones continuas.....	6
Tabla 2.4 Requerimiento 4 – Diseño del módulo de reportes	7
Tabla 2.5 Requerimiento 5 – Programación del módulo: prueba de bondad de ajuste Ji-Cuadrada	7
Tabla 2.6 Requerimiento 6 – Programación del módulo: prueba de bondad de ajuste Kolmogorov-Smirnov	8
Tabla 2.7 Requerimiento 7 – Programación del módulo: prueba de bondad de ajuste Anderson-Darling.....	8
Tabla 2.8 Popularidad de los lenguajes de programación. Índice TIOBE abril 2019	10
Tabla 2.9 Valores de frontera.....	23
Tabla 2.10 Barra de herramientas de histograma	29
Tabla 3.1. Resumen de distribuciones discretas (elaboración propia)	32
Tabla 3.2 Resumen de distribuciones continuas	36
Tabla 4.1 Estimadores de parámetros de algunas distribuciones	42
Tabla 4.2 Valores críticos prueba Ji cuadrada.	44
Tabla 4.3 Valores críticos. Prueba Kolmogorov-Smirnov. (Frías, M.P., 2018)	46
Tabla 4.4 Valores críticos. Prueba Anderson-Darling.....	48
Tabla 5.1 Aforos en Puente Internacional Nuevo Laredo (2015).....	49

Tabla 5.2 Detalle de la prueba Ji-Cuadrada. Datos de aforos.....	50
Tabla 5.3 Minutos entre llegadas de Aeroméxico a CDMX (2013).....	50
Tabla 5.4 Detalle de la prueba K-S. Datos de llegadas de Aeroméxico	52
Tabla 5.5 Pesos del eje motriz. Camiones T3S2 Estación 246/2010	53
Tabla 5.6 Detalle de la prueba A-D. Datos de peso del eje motriz T3S2	54

Sinopsis

Este trabajo presenta el desarrollo de una aplicación informática para hacer pruebas de bondad de ajuste de datos en línea. Las pruebas de bondad de ajuste son pruebas estadísticas de hipótesis que permiten decidir si una muestra de datos determinada puede representarse adecuadamente por alguna ley de probabilidad elegida. Estas pruebas se utilizan con frecuencia en trabajos de investigación y desarrollo tecnológico en el campo del transporte y en diversas ingenierías, para obtener un modelo probabilístico apropiado para datos reales con comportamiento aleatorio.

ABAL (Aplicación para Bondad de Ajuste en Línea) realiza estas pruebas de bondad de ajuste usando tres técnicas comunes para este propósito: Ji-Cuadrada, Kolmogorov-Smirnov y Anderson-Darling, según sean pertinentes para el tipo de datos de entrada.

ABAL se desarrolló en lenguaje Python por su gran versatilidad en la programación y su amplia colección de librerías con rutinas estadísticas ya muy consolidadas y de uso sencillo. El código de programación de ABAL se complementó con la generación de un reporte completo de los resultados de la prueba estadística y su histograma. Esta aplicación se encuentra disponible en la dirección <https://abal.imt.mx/>

Abstract

This work presents the development of a computer application to perform data goodness of fit online. Goodness of fit tests are statistical hypothesis test allowing to decide if a given sample of data may be adequately represented by a some chosen probability law. Goodness of fit tests are frequently used in research and technological development in the field of transport and in various engineering specialties to obtain an appropriate probabilistic model for real data with random behavior.

The online application ABAL (computer Application for Online Goodness of Fit) performs goodness of fit test using three common techniques for this purpose: Chi-square, Kolmogorov-Smirnov and Anderson-Darling, according to the data type of entry.

The ABAL development used Python in view of its great versatility in programming and because it has a collection of libraries with statistical routines that are already well consolidated and easy to use. The ABAL code was supplemented with a complete report of results of the statistical test along with histograms. This application is available at the address <https://abal.imt.mx/>

Resumen ejecutivo

Este trabajo presenta el desarrollo de una aplicación informática en línea que realiza pruebas de bondad de ajuste para los datos de entrada de los usuarios.

Las pruebas de bondad de ajuste son pruebas estadísticas de hipótesis que sirven para decidir si un modelo probabilístico propuesto es adecuado para un conjunto de datos de interés para el usuario. Estas pruebas se requieren con frecuencia en la práctica de investigación y desarrollo en el campo del transporte y en varias especialidades de la ingeniería para tratar adecuadamente a conjuntos de datos reales que tienen comportamiento aleatorio. Las pruebas de bondad de ajuste más comunes que se utilizan en la práctica son Ji-Cuadrada, Kolmogorov-Smirnov y Anderson-Darling.

Ante un conjunto de datos de naturaleza aleatoria, los investigadores o ingenieros que desarrollan proyectos tienen necesidad de establecer un modelo probabilístico adecuado para sus datos, para así poder estimar probabilidades de eventos relevantes para sus proyectos y también para estimar valores esperados de las variables observadas que les permitan dimensionar y prever acciones necesarias según se requieran en los proyectos.

Si un usuario con datos de tipo aleatorio tiene acceso a algún paquete estadístico comercial (el cual implica en algún momento el costo de compra) como Minitab o JMP, podría encontrar en ellos la prueba de bondad de ajuste que requiere. Sin embargo, no todos los paquetes ofrecen las mismas distribuciones de probabilidad para ajustar, y tampoco ofrecen todas las técnicas que el usuario quisiera probar para sus datos. Además, no siempre los usuarios encuentran en los menús de paquetes estadísticos comerciales, el modelo de probabilidad que les interesa.

Si se conocen los procedimientos para estimar los parámetros del modelo probabilístico que se desea probar para los datos, y se conocen las técnicas de bondad de ajuste, los usuarios podrían desarrollar sus propios códigos de computadora o implementarlos en archivos de Excel, lo cual implica dedicar tiempo a probar las rutinas y verificar que se obtengan resultados adecuados.

El desarrollo de la aplicación informática ABAL (Aplicación para Bondad de Ajuste en Línea) que se presenta en este informe resuelve de manera práctica y eficiente la realización de pruebas de bondad de ajuste, generando un reporte de resultados para el usuario en el que se pueden agregar complementos gráficos con histogramas de los datos de entrada y curvas de densidad de la distribución examinada, dando así un reporte más completo.

En el capítulo 1 “Antecedentes” se refiere el origen de este desarrollo de la aplicación ABAL, como una etapa subsecuente de la Publicación Técnica No. 545 “*Modelos de probabilidad en transporte e ingeniería. Usos comunes y ajuste de datos*” en la que se hizo una revisión de las distribuciones de probabilidad más comunes utilizadas en el campo del transporte y diversas especialidades de ingeniería, así como de las aplicaciones exitosas reportadas en la literatura.

El capítulo 2, “Desarrollo de la aplicación informática”, se describen los requerimientos, diagramas de modelado y actividades necesarias para el desarrollo de la aplicación ABAL, todo ello con base en la metodología de desarrollo de software el Modelo Iterativo Incremental. Así mismo se justifica y describen los paquetes de software y lenguajes de programación que conformaron el entorno de desarrollo de la aplicación. Finalizando con la documentación técnica de la primera versión de ABAL 1.0Beta

En el capítulo 3, “Distribuciones de probabilidad en transporte e ingeniería” se hace una síntesis de las principales distribuciones de probabilidad que suelen aparecer en la práctica del transporte y las ingenierías, mostrando sus definiciones de función de probabilidad o de densidad, según sean discretas o continuas, sus formulaciones con los parámetros que las caracterizan y sus fórmulas para media y varianza.

Esta información es básica para que el usuario empiece a verificar la conveniencia de la ley de probabilidad que está probando, ya sea observando la forma del histograma de los datos, o en caso de tener una propuesta de parámetros para el modelo probabilístico, examinar diferencias entre los valores teóricos de media y varianza que resultaran con los efectivamente calculados con los datos de la muestra.

Se agregan también ejemplos típicos de funciones de probabilidad discretas y de funciones de densidad continuas que permiten dar idea del comportamiento del modelo probabilístico ante cambios en los parámetros que definen las distribuciones.

En el capítulo 4 “Técnicas de Bondad de Ajuste” se describen las tres técnicas de ajuste Ji-Cuadrada, Kolmogorov-Smirnov y Anderson-Darling, así como los procedimientos para estimar los parámetros de las distribuciones que se desean probar para los datos de la muestra.

Entre los cálculos requeridos para las pruebas de bondad de ajuste, se necesita estimar frecuencias esperadas de ocurrencia de datos o probabilidades acumuladas esperadas de los datos de la muestra, para compararlas con las que resultarían de ser adecuado el modelo probabilístico que se propone para los datos. Estas comparaciones son la base de las técnicas de bondad de ajuste, en el caso de la Ji-Cuadrada comparando frecuencia observada contra frecuencia esperada, y en el caso Kolmogorov-Smirnov y Anderson-Darling comparando

probabilidad acumulada esperada contra la correspondiente observada en los datos.

Las pruebas de bondad de ajuste utilizan tablas de valores críticos para decidir sobre la calidad del ajuste de los datos examinados. En este capítulo también se muestran las tablas para Ji-Cuadrada, para Kolmogorov-Smirnov y para Anderson-Darling, para que el lector compare los resultados que obtenga de las pruebas.

En el capítulo 5, “Ejemplos de uso de la aplicación con datos reales” se dan varios ejemplos de uso de las técnicas de bondad de ajuste, utilizando datos reales colectados en diversas situaciones relacionadas con problemas de transporte. Se muestran los cálculos realizados paso a paso siguiendo la metodología indicada en el capítulo 4.

El capítulo 6 “Resultados y Conclusiones” resume el desarrollo de la aplicación ABAL que parte de la necesidad de obtener un modelo probabilístico sustentado estadísticamente para los datos de carácter aleatorio que se presentan en el transporte y en las ingenierías, indicando además las características de ABAL y los productos que entrega a sus usuarios. Conforme los usuarios utilicen ABAL en sus proyectos de investigación y desarrollo tecnológico, se podrá obtener retroalimentación de sus opiniones para mejoras futuras a esta primera versión beta de la aplicación.

1 Antecedentes

A finales del primer trimestre de 2019 la Coordinación de Transporte Integrado y Logística del IMT liberó la Publicación Técnica No. 545 “*Modelos de probabilidad en transporte e ingeniería. Usos comunes y ajuste de datos*” en el que se hizo una revisión amplia de los principales usos y aplicaciones reportadas en la literatura, de las distribuciones de probabilidad de uso más común en ingeniería y transporte.

Adicionalmente, el proyecto documenta y resume las técnicas estadísticas de ajuste de datos que se requieren para tener un modelo de probabilidad adecuado para los datos de interés.

En la discusión de las técnicas de bondad de ajuste y su uso práctico, el proyecto TI-02/18 muestra una serie de ejemplos numéricos de aplicación desarrollados con las rutinas estadísticas disponibles en Excel. Varios de estos ejemplos se desarrollaron con datos reales colectados en diversos proyectos de investigación del IMT.

Sin embargo, para que el lector pueda aplicar estas técnicas es necesario las correspondientes rutinas ya sea en una hoja de Excel o en el lenguaje de programación que el usuario.

Este trabajo muestra el desarrollo de una aplicación informática disponible en línea en el Laboratorio Nacional en Sistemas de Transporte y Logística [SiT-LOG Lab] con sede en el IMT, y cuyo acceso está en el sitio web <http://lab-nacional-logistica.imt.mx/> como servicio de consulta técnica estadística para los usuarios de la aplicación.

Si bien en la práctica ingenieril es posible tener acceso a algún paquete comercial de estadística, como por ejemplo MiniTab, JMP o SPSS, las opciones de ajuste de distribuciones no suelen ser las mismas para todos los paquetes, ni todas las distribuciones que pudieran requerirse aparecen en los menús de estos paquetes. Es por ello que en esta aplicación informática se hace énfasis en las distribuciones de probabilidad que son de uso más común en proyectos de investigación y de solución de problemas en ingeniería y en transporte.

Una ventaja adicional de esta aplicación es que se encuentra en línea y es de libre acceso para cualquier usuario; además de que la aplicación se estará revisando y actualizando con la información retroalimentada por los usuarios a fin de mejorar su desempeño y examinar las sugerencias para desarrollos futuros.

Finalmente, para reforzar la información que recibe el usuario de esta aplicación, se tendrá dentro de la aplicación una colección de ejemplos resueltos con datos

reales colectados en diversos trabajos del IMT, con acceso a estos datos reales a fin de que los usuarios comprueben el uso de la aplicación.

En la versión estable de esta aplicación se tendrá acceso en un ambiente amigable para el usuario. Dicha aplicación fue desarrollada en un contexto de software libre, basada en un entorno científico estándar de Anaconda Enterprise Python.

Esta tendencia de utilizar software libre ha permitido avanzar en el IMT en el diseño de aplicaciones de interés práctico, como es el que nos ocupa con las técnicas de bondad de ajuste, para disponer de herramientas útiles en los proyectos de investigación, y poder competir con el software comercial que, aunque ofrece productos de calidad, su adquisición implica altos costos, que particularmente en los institutos de investigación sin apoyo de las grandes empresas, resulta una limitante.

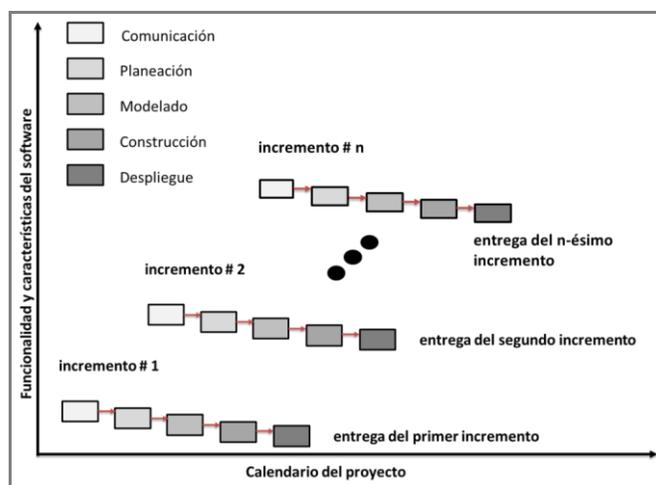
2 Desarrollo de la aplicación informática

Para elaborar la aplicación se tomaron como referencia las prácticas de diseño iterativo y programación basada en objetos, de acuerdo a la metodología iterativa de desarrollo de software. Esta práctica resulta conveniente, ya que puede adaptarse para satisfacer las necesidades específicas del proyecto, y de acuerdo a la Metodología para el Desarrollo de Aplicaciones (MDA) de la SCT, el modelo iterativo es la base del proceso de ingeniería de software para proyectos de este tipo.

2.1 El modelo iterativo incremental

El Modelo Iterativo Incremental permite generar pequeñas versiones sucesivas de la aplicación que aportan incrementos con funcionalidad para ser evaluados. Estos incrementos proporcionan retroalimentación al diseño y guían los cambios de la versión en turno hacia la siguiente etapa con la finalidad de mejorar la cobertura de necesidades de la aplicación y la entrega de características esperadas.

En otras palabras, cada iteración ofrece al usuario un subconjunto de características y funcionalidades a probar que volverán más compleja la aplicación, hasta terminar con el producto final. El esquema de la figura 2.1 ilustra las cinco fases estructurales del Modelo Iterativo Incremental (MII), que definen los flujos de trabajo entre los elementos del proceso de desarrollo.



Fuente: Ingeniería del software. Roger S. Pressman (2010)

Figura 2.1 Modelo Iterativo Incremental

2.2 Fases estructurales

Las fases de: I) Comunicación, II) Planeación, III) Modelado, IV) Construcción y V) Despliegue del MII, permitieron controlar el flujo de trabajo de la aplicación web.

A partir de la definición del objetivo, alcances y requerimientos, se crearon los modelos UML y/o diagramas que, con la combinación de código y pruebas, se logró obtener la publicación de la aplicación en el sitio web del *SiT-LOG Lab*.

A continuación, se describe el desarrollo de la aplicación web en el orden de las fases propuestas por el MII.

2.2.1 Fase I. Comunicación

En la fase inicial de comunicación, se determinó el objetivo, alcances y requerimientos previstos para este trabajo, los cuales se mencionan a continuación.

2.2.1.1 Objetivo

Desarrollar una aplicación estadística en línea para realizar pruebas de bondad de ajuste sobre muestras de datos proporcionadas por el usuario.

2.2.1.2 Alcances

- 1) Desarrollar una aplicación en línea para realizar pruebas de bondad de ajuste a datos aportados por los usuarios.
- 2) Desarrollar los histogramas correspondientes a las distribuciones de probabilidad con parámetros definidos por el usuario.
- 3) Desarrollar un módulo de reportes de las pruebas de bondad de ajuste realizadas con la aplicación.
- 4) Elaborar un informe de investigación con la documentación del sistema computacional que dé soporte a la aplicación, y un manual de consulta rápida para los usuarios.

2.2.1.3 Requerimientos

De acuerdo a Pressman (2010), existen diferentes tipos de requerimientos, sin embargo, por el alcance y objetivo de este trabajo, los tipos de requisitos que definieron la funcionalidad de la aplicación web son: 1) requisitos de las interfaces, 2) requisitos funcionales, y 3) requisitos de mantenimiento. A continuación, se describen cada uno de estos:

Requerimientos de las interfaces

Estos requisitos permiten identificar las interfaces que deben crearse y obtener la relación que existe entre cada una de estas. Así mismo, se describen los elementos que deben proveer para lograr la interacción con el usuario.

Tabla 2.1 Requerimiento 1. Diseño de la interfaz de la página principal.

Número de requisito	RIU-1	
Nombre de requisito	Diseñar y desarrollar la página principal	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	Equipo de trabajo	
Descripción del requisito	Diseñar la página de inicio con los siguientes elementos:	
<p>1.1 Indicar las pruebas de bondad de ajuste: <i>Ji-Cuadrada</i> (Ji-C), <i>Kolmogorov-Smirnov</i> (K-S) y <i>Anderson Darling</i> (A-D).</p> <p>1.2 Indicar los tipos de distribuciones de probabilidad: discretas y continuas.</p> <p>1.3 Indicar las siguientes opciones de cálculo en la distribución discreta a través de la prueba de bondad de ajuste Ji-C:</p> <ul style="list-style-type: none"> • Binomial • Geométrica. Con opción a fallas y pruebas. • Poisson <p>1.4 Indicar las siguientes opciones de cálculo en la distribución continua:</p> <ul style="list-style-type: none"> • Uniforme. Se verifica con las pruebas Ji-C y K-S. • Normal. Se verifica con Ji-C, K-S y A-D. • Exponencial. Se verifica con Ji-C, K-S y A-D. • Lognormal. Se verifica con Ji-C y K-S. • Gamma. Se verifica con Ji-C y K-S. • Weibull. Se verifica con Ji-C y K-S. 		
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial	<input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/ Opcional

Fuente: Elaborado por el IMT

Tabla 2.2 Requerimiento 2 - Diseño de las interfaces para el ingreso de datos en la opción de distribuciones discretas.

Número de requisito	RIU-2	
Nombre de requisito	Diseño de las interfaces para el ingreso de datos.	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	Equipo de trabajo	
Descripción del requisito	Requerir los siguientes elementos en cada una de las interfaces de distribución de probabilidad discretas:	
<p>2.1 Distribución Binomial.</p> <ul style="list-style-type: none"> • Un área para el ingreso de la muestra • Un cuadro de texto de tipo numérico para el ingreso del número de ensayos. • Cuadros de texto para ingresar el título del histograma, así como para el eje "x". • Un botón para activar el cálculo. • Una sección para mostrar los resultados del cálculo y otra para mostrar el histograma. <p>2.2 Distribución Geométrica fallas o ensayos:</p> <ul style="list-style-type: none"> • Un área para el ingreso de la muestra. • Cuadros de texto para ingresar el título del histograma, así como para el eje "x". • Un botón para activar el cálculo. • Una sección para mostrar el resultado y otra para mostrar el histograma. 		

2.3 Distribución Poisson: <ul style="list-style-type: none"> • Un área para el ingreso de la muestra. • Un botón para activar el cálculo. • Una sección para mostrar el resultado y otra para mostrar su histograma. 			
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial	<input type="checkbox"/> Media/Deseado	<input type="checkbox"/> Baja/ Opcional

Fuente: Elaborado por el IMT

Tabla 2.3 Requerimiento 3 - Diseño de las interfaces para el ingreso de datos en la opción de distribuciones continuas.

Número de requisito	RIU-3		
Nombre de requisito	Diseño de las interfaces para el ingreso de datos.		
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción	
Fuente del requisito	Equipo de trabajo		
Descripción del requisito	Requerir los siguientes elementos en cada una de las interfaces de distribución de probabilidad continuas:		
3.1 Distribución Uniforme: <ul style="list-style-type: none"> • Un área para el ingreso de la muestra. • Cuadros de texto para ingresar el título del histograma, así como para el eje "x". • Un botón para activar el cálculo. • Una sección para mostrar el resultado y otra para mostrar el histograma. 			
3.2 Distribución Normal. <ul style="list-style-type: none"> • Un área para el ingreso de la muestra • Cuadros de texto para ingresar el título del histograma, así como para el eje "x". • Un botón para activar el cálculo. • Una sección para mostrar el resultado y otra para mostrar el histograma. 			
3.3 Distribución Exponencial: <ul style="list-style-type: none"> • Un área para el ingreso de la muestra. • Cuadros de texto para ingresar el título del histograma, así como para el eje "x". • Un botón para activar el cálculo. • Una sección para mostrar el resultado y otra para mostrar el histograma. 			
3.4 Distribución Lognormal: <ul style="list-style-type: none"> • Un área para el ingreso de la muestra. • Cuadros de texto para ingresar el título del histograma, así como para el eje "x". • Un botón para activar el cálculo. • Una sección para mostrar el resultado y otra para mostrar el histograma. 			
3.5 Distribución Weibull: <ul style="list-style-type: none"> • Un área para el ingreso de la muestra. • Cuadros de texto para ingresar el título del histograma, así como para el eje "x". • Un botón para activar el cálculo. • Una sección para mostrar el resultado y otra para mostrar el histograma. 			
3.6 Distribución Gamma: <ul style="list-style-type: none"> • Un área para el ingreso de la muestra. • Cuadros de texto para ingresar el título del histograma, así como para el eje "x". • Un botón para activar el cálculo. 			

• Una sección para mostrar el resultado y otra para mostrar el histograma.			
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial	<input type="checkbox"/> Media/Deseado	<input type="checkbox"/> Baja/Opcional

Fuente: Elaborado por el IMT

Tabla 2.4 Requerimiento 4 – Diseño del módulo de reportes

Número de requisito	RIU-4		
Nombre de requisito	Diseño del módulo de reportes.		
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción	
Fuente del requisito	Equipo de trabajo		
Descripción del requisito	Todas las interfaces de distribución deberán tener un módulo de reportes, el cual mostrará los resultados del cálculo realizado. Sus características son:		
	4.1 Activar un botón llamado “Descargar PDF”.		
	4.2 Esta acción deberá mostrar el histograma en un documento con extensión pdf, así como los resultados del cálculo realizado.		
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial	<input type="checkbox"/> Media/Deseado	<input type="checkbox"/> Baja/Opcional

Fuente: Elaborado por el IMT

Requerimientos funcionales

A continuación, se presentan las funcionalidades internas que debe procesar la aplicación para poder cumplir con el cálculo de las pruebas de bondad de ajuste: Ji-Cuadrada, Kolmogorov-Smirnov y Anderson-Darling.

Tabla 2.5 Requerimiento 5 – Programación del módulo: prueba de bondad de ajuste Ji-Cuadrada

Número de requisito	RIU-5		
Nombre de requisito	Programación del módulo: prueba de bondad de ajuste Ji-Cuadrada		
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción	
Fuente del requisito	Equipo de trabajo		
Descripción del requisito	Este módulo servirá como base para el cálculo de las distribuciones de probabilidad que así lo requieran. Para su obtención se deberán programar las siguientes funciones:		
	5.1 Validar la muestra ingresada.		
	5.2 Realizar una función que calcule los parámetros iniciales para la distribución teórica, los cuales servirán como base para el procesamiento de los cálculos principales. Estos son: número de intervalos, mínimo, máximo, promedio, varianza de la muestra, varianza de la población, rango, amplitud, número de elementos, desviación estándar de la muestra y población.		
	5.3 Construir la tabla de frecuencias y calcular la frecuencia observada.		
	5.4 Estimar los parámetros para la distribución teórica que se propone.		
	5.5 Calcular la frecuencia esperada con la ley de probabilidad teórica.		
	5.6 Calcular los valores Ji-Cuadrada y determinar si la prueba es significativa.		
Prioridad del requisito	<input checked="" type="checkbox"/> Alta/Esencial	<input type="checkbox"/> Media/Deseado	<input type="checkbox"/> Baja/Opcional

Fuente: Elaborado por el IMT

Tabla 2.6 Requerimiento 6 – Programación del módulo: prueba de bondad de ajuste Kolmogorov-Smirnov

Número de requisito	RIU-6	
Nombre de requisito	Programación del módulo: prueba de bondad de ajuste kolmogorov-Smirnov	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	Equipo de trabajo	
Descripción del requisito	Este requisito servirá como base para el cálculo de las distribuciones de probabilidad que así lo requieran. Para su obtención se deberán programar las siguientes funciones:	
<p>6.1 Validar la muestra ingresada.</p> <p>6.2 Realizar una función que calcule los parámetros iniciales para la distribución teórica, los cuales servirán como base para el procesamiento de los cálculos principales. Estos son: número de intervalos, mínimo, máximo, promedio, varianza de la muestra, varianza de la población, rango, amplitud, número de elementos, desviación estándar de la muestra y la población.</p> <p>6.3 Construir la tabla de frecuencias y calcular la frecuencia observada.</p> <p>6.4 Estimar los parámetros para la distribución teórica que se propone.</p> <p>6.5 Calcular la frecuencia esperada con la ley de probabilidad teórica.</p> <p>6.6 Calcular los valores Kolmogorov-Smirnov y determinar si la prueba es significativa.</p>		

Fuente: Elaborado por el IMT

Tabla 2.7 Requerimiento 7 – Programación del módulo: prueba de bondad de ajuste Anderson-Darling

Número de requisito	RIU-7	
Nombre de requisito	Programación del módulo: prueba de bondad de ajuste Anderson-Darling	
Tipo	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
Fuente del requisito	Equipo de trabajo	
Descripción del requisito	Este requisito servirá como base para el cálculo de las distribuciones de probabilidad que así lo requieran. Para su obtención se deberán programar las siguientes funciones:	
<p>7.1 Validar la muestra ingresada.</p> <p>7.2 Realizar una función que calcule los parámetros iniciales para la distribución teórica, los cuales servirán como base para el procesamiento de los cálculos principales. Estos son: número de intervalos, mínimo, máximo, promedio, varianza de la muestra, varianza de la población, rango, amplitud, número de elementos, desviación estándar de la muestra y la población.</p> <p>7.3 Construir la tabla de frecuencias y calcular la frecuencia observada.</p> <p>7.4 Estimar los parámetros para la distribución teórica que se propone.</p> <p>7.5 Calcular la frecuencia esperada con la ley de probabilidad teórica.</p> <p>7.6 Calcular los valores Anderson-Darling y determinar si la prueba es significativa.</p>		

Fuente: Elaborado por el IMT

2.2.2 Fase II. Planeación

En esta fase se definieron los recursos tecnológicos necesarios que permitieron cumplir con el desarrollo y publicación de la aplicación. A continuación, se describen los recursos tecnológicos necesarios, los productos por obtener y el plan de actividades.

Recursos técnicos

Para desarrollar este proyecto se utilizaron los recursos tecnológicos disponibles en el IMT los cuales resultaron óptimos para el cumplimiento del objetivo. A continuación, se presentan sus características técnicas.

Equipo de cómputo:

- Equipo con Windows 10 Pro
- Procesador Intel® Core™ i5-7500 CPU @ 3.40GHz 3.41 GHz
- Memoria RAM 8.00 GB
- Sistema operativo de 64bits, procesador x64

Servidor Web:

- Spec
- Procesador: Intel Xeon 2.40GHz, 11M Cache
- Sistema operativo: Windows Server
- Memoria: 16GB
- Disco duro: 500GB 7.2K RPM

Python fue el lenguaje de programación elegido debido a las ventajas que este lenguaje aporta. Algunas de sus ventajas son:

- Cuenta con un amplio soporte de ayuda para la comunidad de programadores.
- Tiene una biblioteca estándar y versátil que se encuentra disponible de inmediato, sin la necesidad de descargar paquetes por separado.
- Es un lenguaje de programación multiparadigma, por ejemplo: orientación a objetos, estructurada y programación imperativa.
- Es multiplataforma, por lo tanto, su código fuente puede ser ejecutable entre varios sistemas operativos.
- Es un lenguaje interpretado¹, no compilado. Usa tipado dinámico² y es fuertemente tipado³.
- Es un lenguaje de código abierto.
- Se encuentra entre los primeros lugares de popularidad. Como se observa en la tabla 2.8, hasta el mes de abril ocupa la posición número cuatro, por arriba del lenguaje R con la posición número dieciséis.

¹ Su código no necesita ser preprocesado mediante un compilador, eso significa que el ordenador es capaz de ejecutar las instrucciones dadas por el programador sin necesidad de leer y traducir exhaustivamente todo el código.

² Significa que los objetos en tiempo de ejecución (valores) tienen un tipo de dato.

³ Significa que el tipo de valor no cambia repentinamente. Cada cambio de tipo requiere una conversión explícita.

Tabla 2.8 Popularidad de los lenguajes de programación. Índice⁴ TIOBE abril 2019

Apr 2019	Apr 2018	Change	Programming Language	Change
1	1	Java	15.035%	-0.74%
2	2	C	14.076%	+0.49%
3	3	C++	8.838%	+1.62%
4	4	Python	8.166%	+2.36%
5	6	Visual Basic .NET	5.795%	+0.85%
6	5	C#	3.515%	-1.75%
7	8	JavaScript	2.507%	-0.99%
8	9	SQL	2.272%	-0.38%
9	7	PHP	2.239%	-1.98%
10	14	Assembly language	1.710%	+0.05%

Fuente: TIOBE. The software quality company.

En este sentido, se puede decir que Python cumplió con las expectativas de este proyecto, por lo que se descargó y usó el entorno de desarrollo científico *Anaconda Enterprise Python*. El cual, proporcionó las herramientas y librerías necesarias para cumplir con los alcances y requisitos previamente descritos en este proyecto.

Dentro de las librerías disponibles, el ecosistema de *Scipy* basado en *Python*, específicamente *NumPy*, *Pandas* y *Scipy.stats*, solventaron la codificación de rutinas de integración y optimización, así como el cálculo de funciones estadísticas de datos numéricos (valores reales).

Para representar el resultado de dichas funciones en gráficos explicativos, particularmente en histogramas, se utilizaron las librerías *Matplotlib* y *Bokeh*.

Como el análisis de datos de este proyecto requirió ser integrado en una aplicación web en un lapso corto, se optó por desarrollar cada página con un

⁴ Las calificaciones se basan en el número de ingenieros calificados en todo el mundo, así como cursos y proveedores externos. Los motores de búsqueda como Google, Bing, Yahoo, Wikipedia, Amazon, YouTube y Baidu se utilizan para calcular sus calificaciones.

diseño minimalista, es decir, reducida solo a lo esencial. El *micro-framework* Flask y el motor de plantillas Jinja2, resultaron ser excelentes para este tipo de diseño.

Así mismo, se usaron otras librerías que, en combinación con los lenguajes de desarrollo web HTML, Ajax y CSS, contribuyeron al desarrollo de la vista de la aplicación. Estas son: bootstrap, bokeh-widgets, bokeh-tables, jquery, jexcel, jspdf y MathJax-master.

Productos por obtener

Como resultado, se tiene la presente publicación y una aplicación informática denominada ABAL 1.0beta de acceso gratuito. Se planea continuar enriqueciendo la aplicación de funcionalidades, por ejemplo, agregando modelos de probabilidad, diferentes tipos de histogramas e incluso combinar métodos que ofrezcan un cálculo más preciso. Y con esto, llegar a obtener un producto final con características y funcionalidades que otros paquetes estadísticos no ofrecen.

2.2.3 Fase III. Modelado

En esta actividad se representaron modelos UML a fin de clarificar los requerimientos de la aplicación y el diseño que los integrará. Estos modelos son: casos de uso de la aplicación, modelo de secuencia de los datos y diagramas de clases.

2.2.3.1 Casos de uso de la aplicación

En los siguientes cuadros se explica el actor que interactúa en los diferentes procesos, así como, la funcionalidad que ofrece la aplicación.

Nombre:	Usuario general de la aplicación
Descripción:	Representa a un usuario que requiere hacer pruebas de bondad de ajuste.

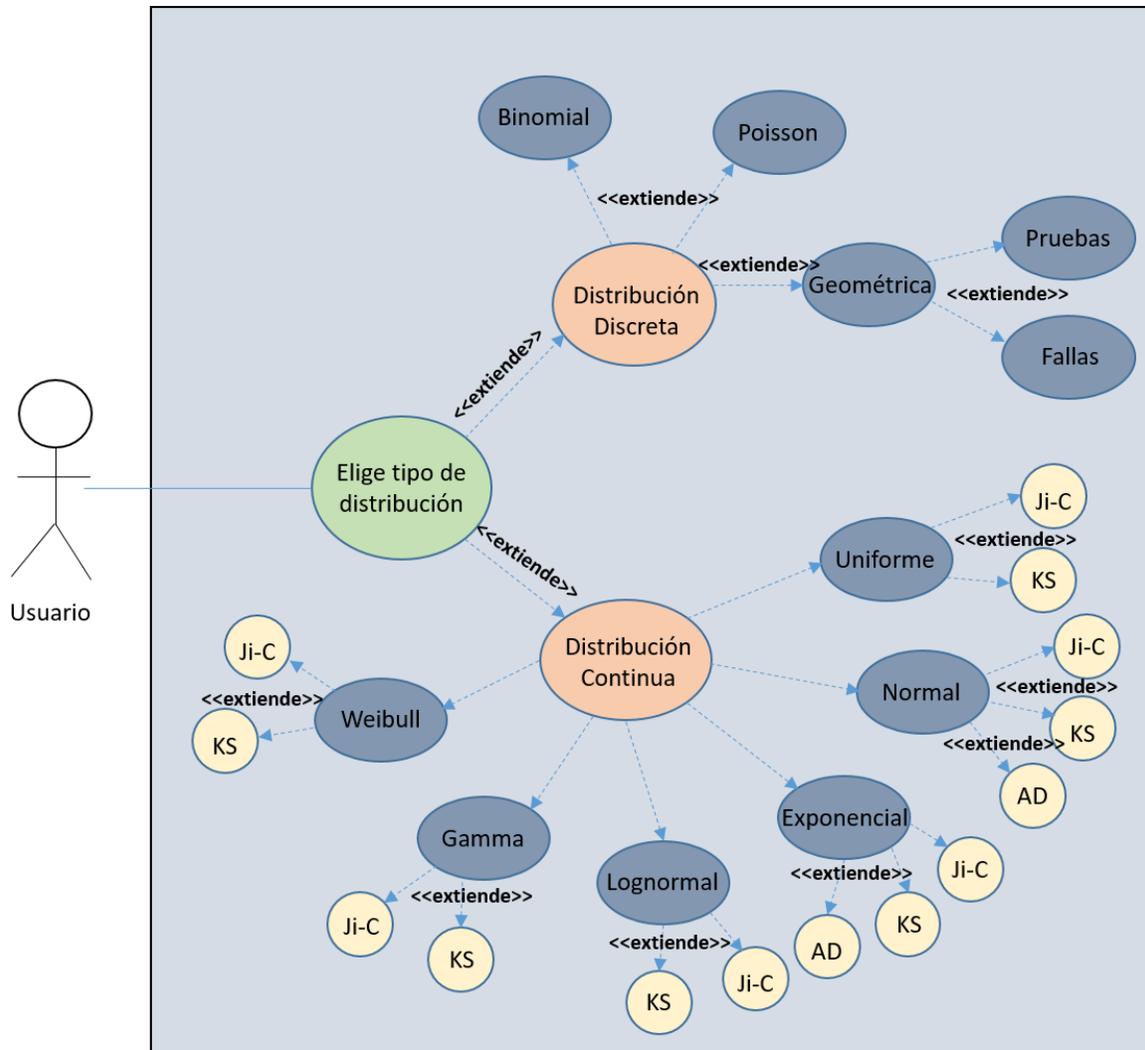
Nombre:	Seleccionar distribución de probabilidad
Descripción:	Permite seleccionar el tipo de distribución. Estas pueden ser: Binomial, Geométrica (fallas y pruebas), Poisson, Uniforme, Normal, Exponencial, Lognormal, Gamma y Weibull.
Actores:	Usuario

Precondiciones: el usuario debe elegir una distribución de probabilidad.
Poscondiciones: la aplicación hará los cálculos correspondientes.

Nombre:	Seleccionar prueba de bondad de ajuste
Descripción: permite seleccionar el tipo de prueba de bondad de ajuste, estas pueden ser: Ji-Cuadrada, Kolmogorov-Smirnov y Anderson-Darling.	
Actores: usuario	
Precondiciones: el usuario debe elegir una prueba de bondad de ajuste.	
Poscondiciones: la aplicación hará los cálculos correspondientes.	

Nombre: Ingresar datos	
Descripción: la aplicación procesará los datos que ingrese el usuario.	
Actores: usuario	
Precondiciones: el usuario deberá ingresar la muestra, los parámetros correspondientes al tipo de distribución, así como, información para graficar el histograma.	
Poscondiciones: la aplicación hará los cálculos correspondientes.	

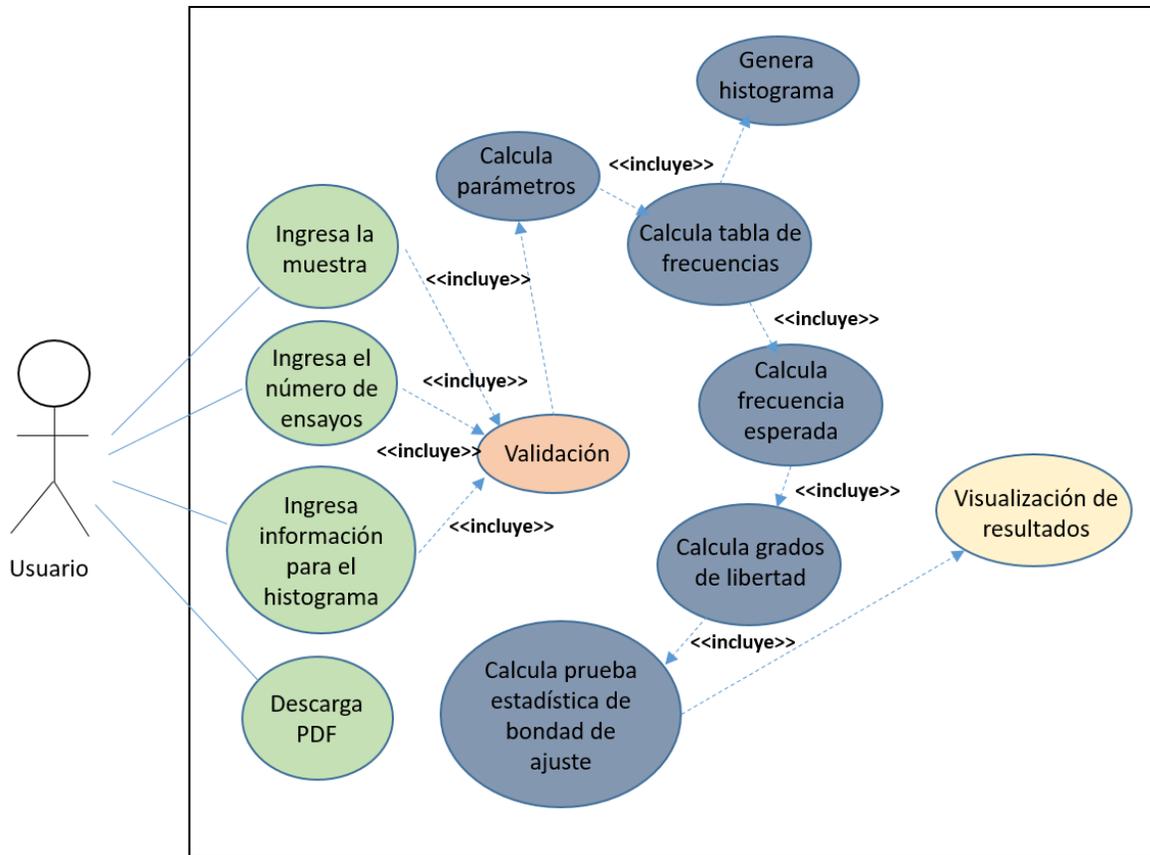
Los siguientes diagramas de casos de uso muestran las acciones y reacciones del comportamiento de la aplicación desde la perspectiva del usuario, desencadenadas en respuesta al estímulo de un actor. El diagrama de la figura 2.2 muestra el proceso donde el usuario debe elegir el tipo de distribución y la prueba de bondad de ajuste a calcular.



Fuente: Elaborado por el IMT

Figura 2.2 Diagrama de caso de uso: pruebas estadísticas de bondad de ajuste.

Posteriormente, para que la aplicación pueda realizar el cálculo correspondiente el usuario deberá ingresar los datos de la muestra, los parámetros de la distribución, así como, los datos del histograma. En la figura 2.4 se muestra un ejemplo de modelado del cálculo de la distribución Binomial.



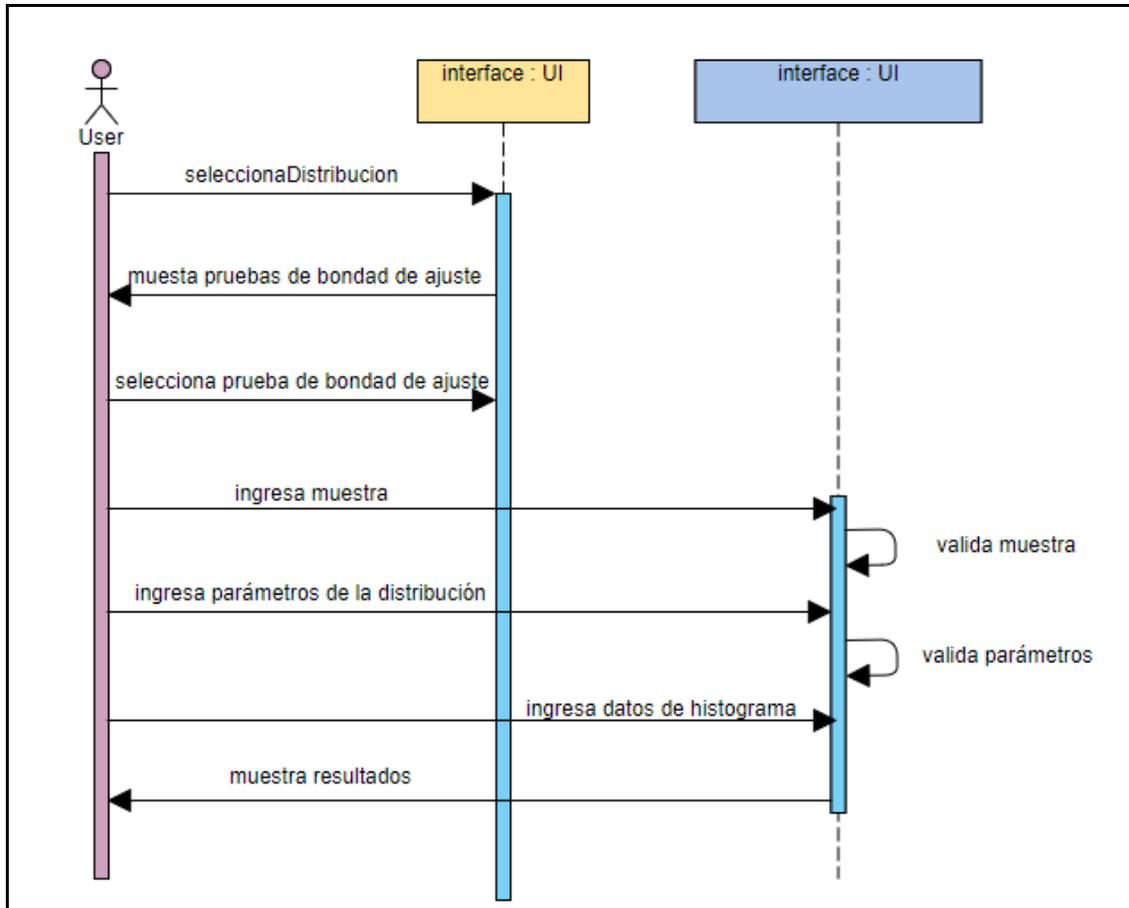
Fuente: Elaborado por el IMT

Figura 2.3 Diagrama de caso de uso: modelado de la distribución binomial.

Para el resto de las distribuciones (Poisson, Geométrica, Uniforme, Normal, etc), el modelo de caso de uso es similar al ejemplo presentado en la figura 2.3 La diferencia radica en los parámetros ingresados por el usuario, en este caso, el número de ensayos.

2.2.3.2 Modelo de secuencia de los datos

A diferencia de los casos de uso, en el siguiente diagrama de secuencia se muestran los elementos de comportamiento del proceso que debe seguir el usuario para calcular la prueba de bondad de ajuste.



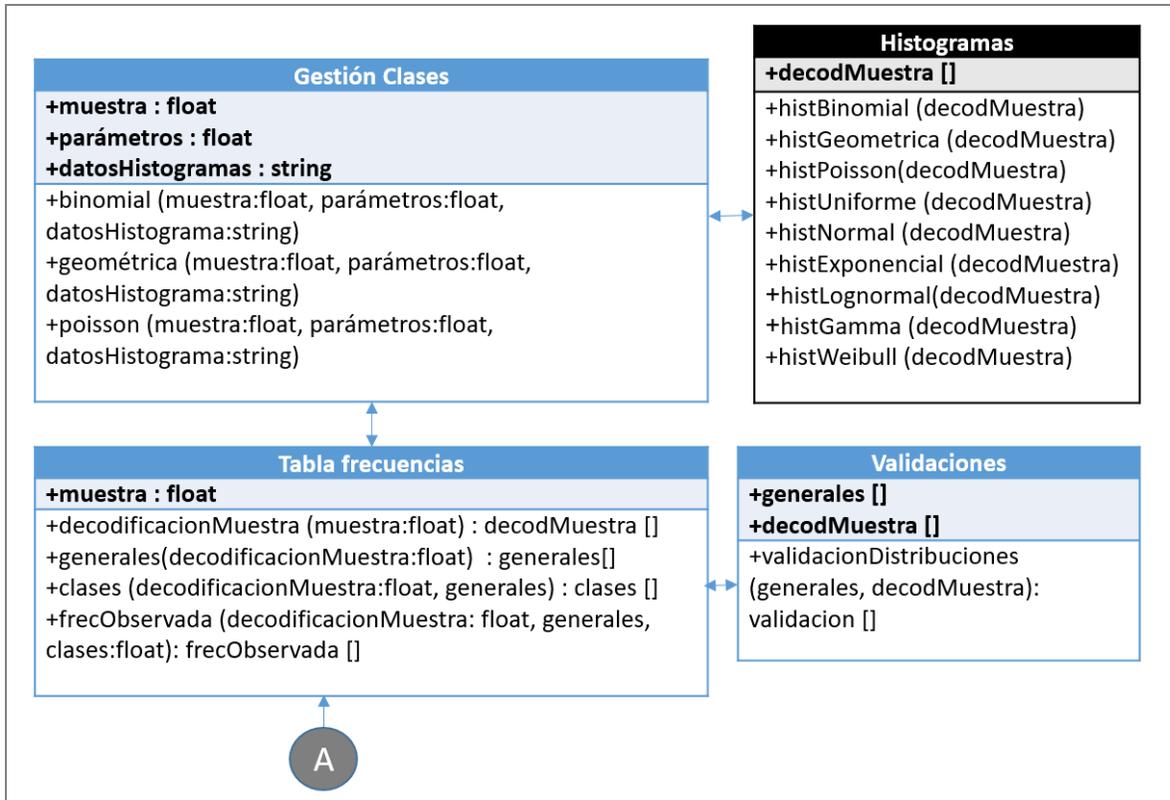
Fuente: Elaborado por el IMT

Figura 2.4 Secuencia de procesamiento de los datos.

2.2.3.3 Diagramas de clases

Los diagramas de clases aportaron una visión estática de la estructura de la aplicación. En las siguientes figuras se muestran los diagramas de clases de las distribuciones discretas y continuas, en las cuales, se modelaron atributos, operaciones, relaciones y asociaciones con otras clases.

Para las distribuciones discretas se modelaron ocho clases. De acuerdo con la figura 2.5 el flujo de información comienza en el módulo “Gestión Clases”, aquí se reciben los parámetros ingresados por los usuarios, posteriormente este módulo se comunica con “Tabla frecuencias”, el cual se encarga de preparar los datos para ser validados y procesados por algunas de las distribuciones disponibles, por ejemplo: “Distribución Binomial” y “Prueba JiCuadrada Binomial”.

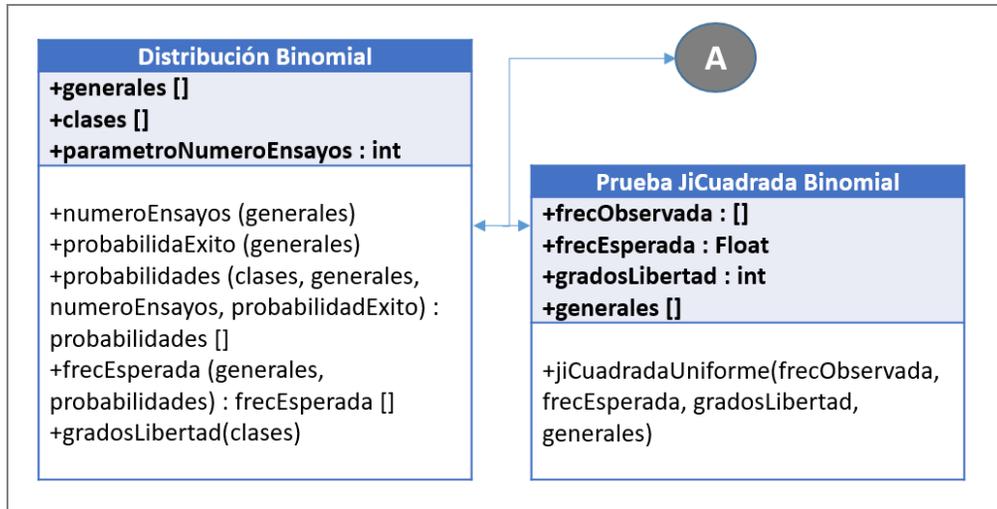


Fuente: Elaborado por el IMT

Figura 2.5 Diagrama de clases para las distribuciones discretas.

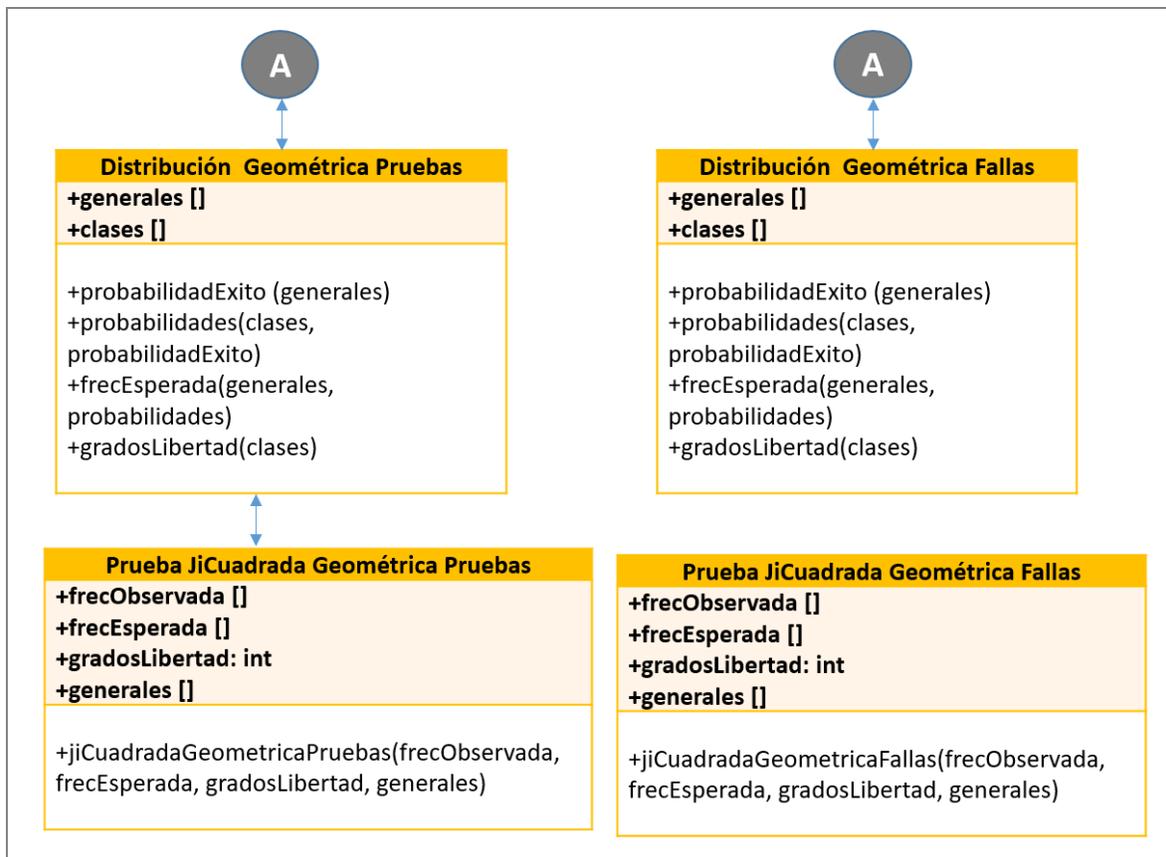
La figura 2.6 muestra las funciones necesarias para la distribución binomial. Por ejemplo, se necesita del cálculo de: número de ensayos, probabilidad de éxito, probabilidades, frecuencia esperada, grados de libertad y por último su prueba JiCuadrada.

Para la distribución geométrica (fallas y pruebas), una vez cumplido el flujo de la figura 2.5, se procede al cálculo de: probabilidad de éxito, probabilidades, frecuencia esperada, grados de libertad y su prueba JiCuadrada como se observa en la figura 2.7.



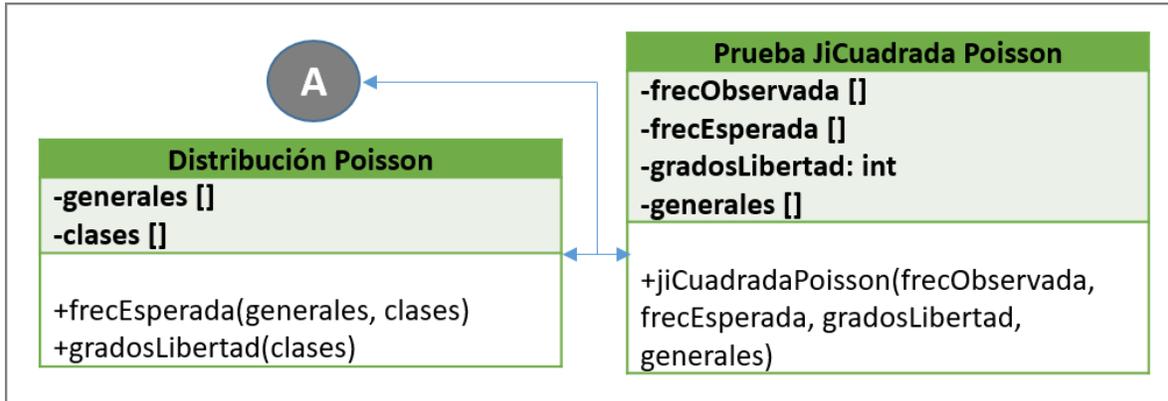
Fuente: Elaborado por el IMT

Figura 2.6 Diagrama de clases distribución binomial



Fuente: Elaborado por el IMT

Figura 2.7 Diagrama de clases distribución Geométrica (pruebas y fallas)

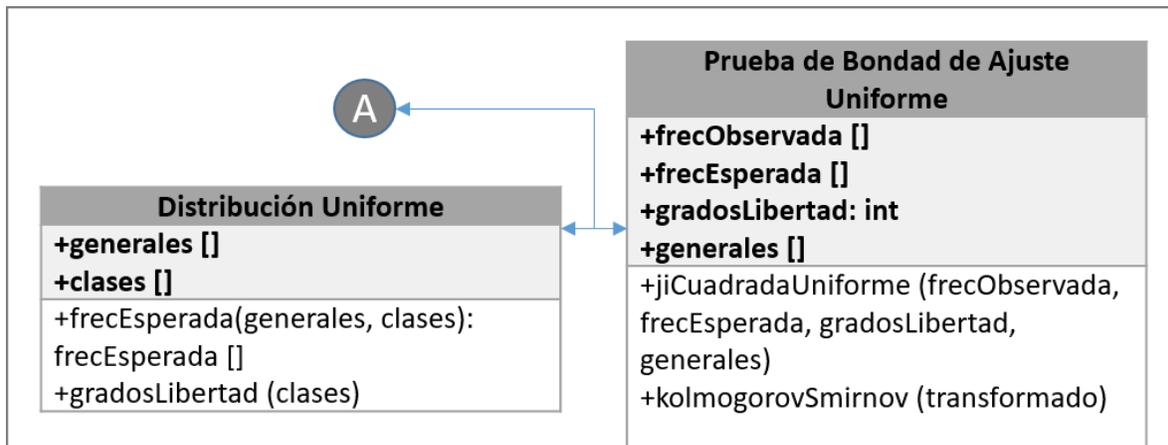


Fuente: Elaborado por el IMT

Figura 2.8 Diagrama de clases distribución Poisson

En las distribuciones continuas se modelaron diez clases. El flujo de información, al igual que las distribuciones discretas, comienza en el módulo “gestión datos”, el cual recibe los parámetros ingresados por el usuario. Después, estos son enviados a la clase “tabla frecuencias” para codificar la muestra y posteriormente ser verificada en la clase “validaciones”.

Si pasa la validación, se construye la tabla de frecuencias que será llamada por alguna de las distribuciones continuas y pruebas de bondad de ajuste, como: Ji-Cuadrada, Kolmogorov-Smirnov y Anderson-Darling.

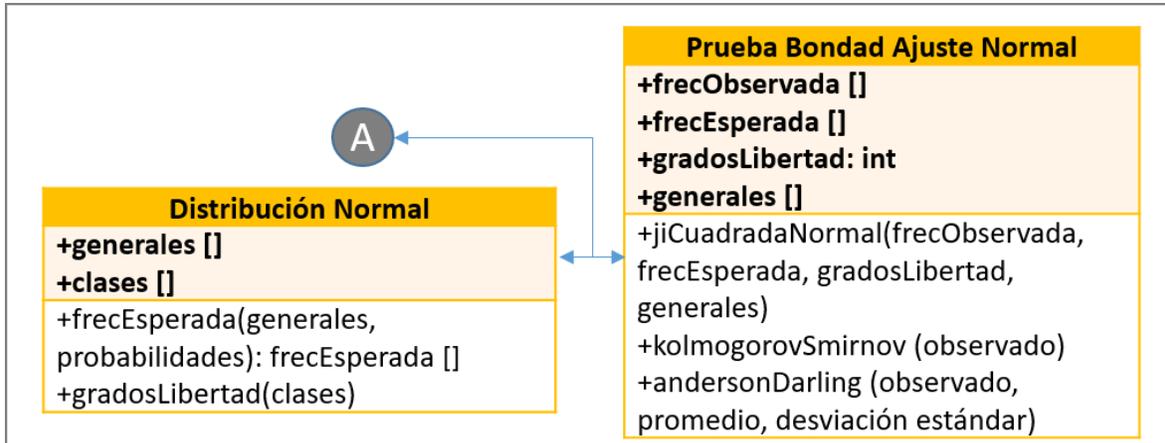


Fuente: Elaborado por el IMT

Figura 2.9 Diagrama de clases distribución Uniforme

En la figura 2.9 se observan las funciones necesarias para el cálculo de la distribución Uniforme, estas son: frecuencia esperada y grados de libertad, con sus respectivas pruebas: JiCuadrada y Kolmogorov

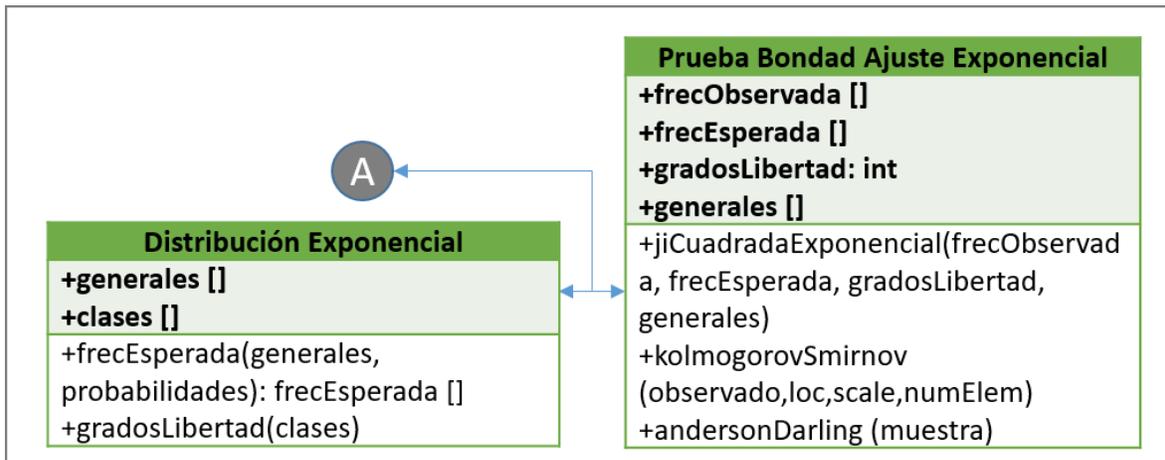
Para el cálculo de la distribución Normal se necesita de las funciones: frecuencia esperada y grados de libertad, además de sus pruebas: JiCuadrada, Kolmogorov y Anderson, como se muestra en la figura 2.10.



Fuente: Elaborado por el IMT

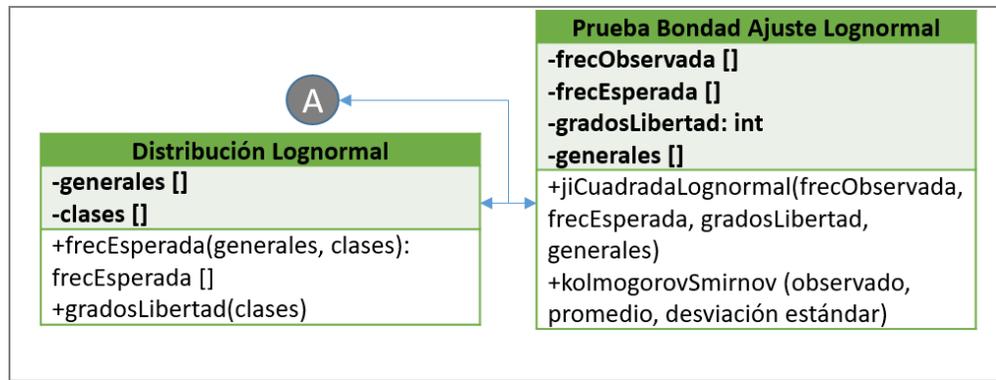
Figura 2.10 Diagrama de clases distribución Normal

De igual forma, en la distribución Exponencial se necesita de las funciones: frecuencia esperada y grados de libertad, además de sus pruebas JiCuadrada, Kolmogorov y Anderson. En dichas pruebas se observan los parámetros necesarios para su cálculo (ver figura 2.11).



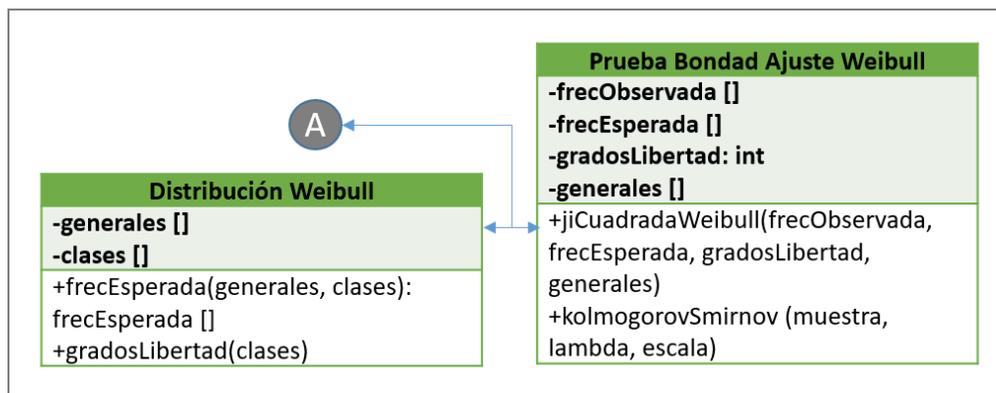
Fuente: Elaborado por el IMT

Figura 2.11 Diagrama de clases distribución Exponencial



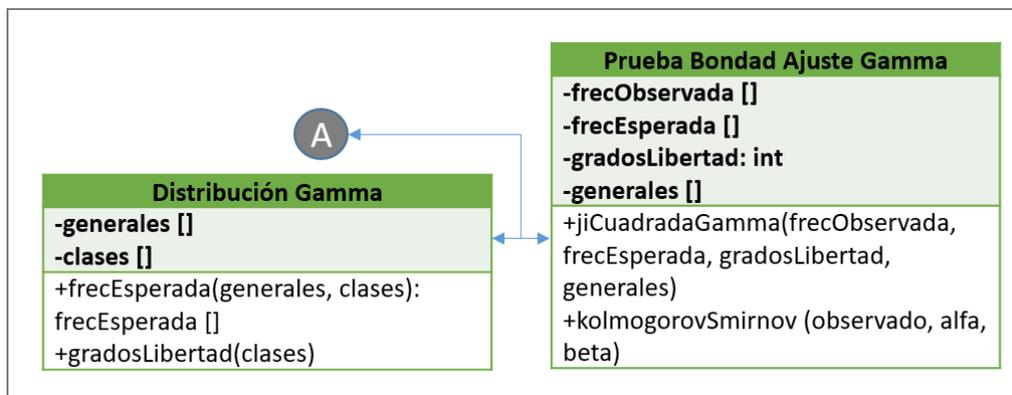
Fuente: Elaborado por el IMT

Figura 2.12 Diagrama de clases distribución Lognormal.



Fuente: Elaborado por el IMT

Figura 2.13 Diagrama de clases distribución Weibull.



Fuente: Elaborado por el IMT

Figura 2.14 Diagrama de clases distribución Gamma

Para las distribuciones Lognormal, Weibull y Gamma se requiere de las funciones: frecuencia esperada y grados de libertad. Y se necesita de ciertos parámetros para el cálculo de la prueba de ajuste, mismos que se muestran en las figuras 2.12, 2.13 y 2.14.

Todas las distribuciones tienen comunicación con la clase “histogramas”, la cual como su nombre lo indica, retornará un gráfico en una imagen que será presentada en la sección de resultados. En total, se desarrollaron veintidós clases para la aplicación.

2.2.4 Fase IV. Construcción

En esta fase se realizaron diversas pruebas durante la construcción de la aplicación. Empezaremos definiendo esta. La prueba de una aplicación web es una colección de actividades relacionadas con una sola meta: descubrir errores en el contenido, función, utilidad, navegabilidad, rendimiento, capacidad y seguridad de esa aplicación (Pressman, 2010).

Para lograr esto, se aplicó una estrategia de prueba que abarcó tanto revisiones como pruebas ejecutables que permitieron detectar y eliminar tantos errores como fue posible antes de poner en línea la aplicación web. En estas pruebas participaron desarrolladores web y matemáticos.

El proceso de prueba comenzó enfocándonos en los aspectos visibles para el usuario de la aplicación y avanzó hacia pruebas que ejercitaron la tecnología y la infraestructura. Esta consistió en siete pasos: pruebas de contenido, pruebas de interfaz, prueba de componente, prueba de configuración, seguridad y rendimiento.

Durante la fase de construcción se realizaron diversas pruebas tanto de lado del cliente como del servidor, con la finalidad de garantizar el correcto funcionamiento de la aplicación. Estas pruebas se pueden clasificar como: de interfaz de usuario, prueba a nivel de componentes, configuración, seguridad y pruebas de rendimiento.

2.2.4.1 Pruebas de contenido

Estas pruebas permitieron identificar:

- Errores sintácticos (por ejemplo, errores tipográficos o gramaticales) en el contenido de las interfaces y representaciones gráficas.
- Errores semánticos, es decir, errores en la completitud de la información en cualquier objeto de contenido que se presente conforme ocurre la navegación.
- Errores en la organización o estructura del contenido que se presenta al usuario final.

2.2.4.2 Pruebas de interfaz de usuario

Aquí se validaron los aspectos estéticos de la interfaz de usuario. La estrategia global para esta prueba fue: descubrir errores relacionados con los mecanismos de interfaz específicos, descubrir errores en la forma como la interfaz implementa la semántica de navegación y la funcionalidad de la aplicación o el despliegue del contenido. Para lograrlo se realizó lo siguiente:

Pruebas de mecanismos de interfaz:

- **Vínculos.** Cada vínculo de navegación se probó para garantizar que se alcanzó el objetivo de contenido o funciones apropiadas.
- **Formularios.** Se revisó que las etiquetas de los formularios identificaran correctamente los campos, y los que se identificaron como obligatorios, se identifiquen visualmente para el usuario. Se validó que el servidor recibiera toda la información contenida en los formularios y que las funciones del navegador (flecha “retroceso”) no corrompan la entrada de datos a los formularios. Y se comprobó que los errores en los datos ingresados funcionarían de manera adecuada, proporcionando mensajes de alerta para el usuario.
- **HTML dinámico.** Se ejecutó el contenido HTML para asegurar el despliegue dinámico correcto. Además, se garantizó la compatibilidad de la aplicación y se realizaron pruebas en el navegador: Chrome.
- **Prueba de la semántica de la interfaz.** Se revisó que las instrucciones presentadas al usuario fueran claras. Los casos de uso definidos previamente sirvieron como secuencia de prueba para descubrir errores que evitaran que un usuario logre el objetivo asociado a cada caso de uso. El objetivo de esto fue determinar si la aplicación proporciona manejo y recuperación efectivos del error.

2.2.4.3 Pruebas en el nivel de componente

También llamada prueba de función, se enfoca en un conjunto de pruebas que intentan descubrir errores en los componentes de software. Estas pruebas comienzan en el ingreso de datos en los formularios precedido de la acción de un botón para iniciar la ejecución. A continuación se explican:

- **Análisis de valor de frontera.** Los datos que fueron ingresados en los formularios se probaron en sus fronteras establecidas. Con ello se validó que los parámetros y la muestra ingresada por el usuario estuvieran dentro del rango establecido. En la tabla 2.9 se presentan los valores de frontera validados para cada distribución estadística.

Tabla 2.9 Valores de frontera

Distribución	Valor de frontera en los parámetros
Binomial	$n > 1; 0 < p < 1; k = 0, 1, 2, \dots, n$
Geométrica (núm. de fallas)	$0 < p < 1; k = 0, 1, 2, \dots$
Geométrica (núm. de ensayos)	$0 < p < 1; k = 1, 2, \dots$
Poisson	$\lambda > 0; k = 0, 1, 2, \dots$
Uniforme	$x = 1, 2, 3, \dots, n$ $-\infty < a, b < \infty; a < b$
Normal	$-\infty < x < \infty$ $-\infty < \mu < \infty; \sigma > 0$
Exponencial	$x \geq 0;$ $\lambda > 0$
Lognormal	$x, \sigma > 0; -\infty < \mu < \infty$
Gamma	$\alpha, \beta > 0; x \geq 0;$
Weibull	$\lambda, k > 0; x > 0$

Fuente: Elaborado por el IMT

2.2.4.4 Prueba de configuración

Esta prueba consiste en probar un conjunto de probables configuraciones en el lado del cliente y servidor con la finalidad de garantizar que la experiencia del usuario será la misma para todos y que aislará los errores que pueden ser específicos de una configuración particular.

Pruebas del lado del servidor

La actividad es, experimentar con diversos casos de prueba de configuración para verificar que la configuración del servidor puede soportar la aplicación. Durante esta prueba se efectuaron las siguientes preguntas que validaron la operación correcta del servidor:

Pruebas de configuración del lado del servidor	
¿La aplicación es compatible con el servidor?	Sí
¿Los archivos de sistema, directorios y datos de sistema relacionados se crean correctamente cuando la aplicación es operativa?	Sí
¿Las medidas de seguridad del sistema, por ejemplo, <i>firewalls</i> , permitieron a la aplicación ejecutarse y atender a los usuarios sin interferencia?	Sí

Pruebas del lado del cliente

Del lado del cliente, las pruebas de configuración se enfocaron en la compatibilidad de la aplicación con las configuraciones de los siguientes componentes:

- Hardware: CPU y memoria.
- Sistemas operativos: Microsoft Windows y un sistema operativo móvil.
- Software navegador: Chrome.

2.2.4.5 Pruebas de seguridad

Debido a que la aplicación desarrollada es simplemente una herramienta de cálculo estadístico, no representa un riesgo ante el robo de información, sin embargo, los formularios pueden presentar un riesgo al estar expuestos para la transmisión de comandos no autorizados

Para ello se habilitó la protección CSRF⁵ (Cross-site request forgery) disponible para aplicaciones desarrolladas en Flask.

```
from flask_wtf.csrf import CSRFProtect
csrf = CSRFProtect(app)
```

Además se utilizó el campo `form.csrf_token` de *Flask-WTF* en todos los formularios de la aplicación web.

```
<form action="/user" method="POST">
  {{ form.csrf_token }}
  {% if form.csrf_token.errors %}
    <div class="warning">You have submitted an invalid CSRF token</div>
  {% endif %}
```

2.2.4.6 Pruebas de rendimiento

En este tipo de pruebas no se detectaron problemas de rendimiento que pudieron ser originados por falta de recursos en el lado del servidor, sin embargo, se observó que, en muestras probadas con datos del orden de millones, la aplicación tarda aproximadamente 3 minutos en procesar el resultado, lo cual es inadmisibles para una aplicación práctica.

⁵ Solicitud de falsificación entre sitios, también conocida como ataque con un solo clic, es un tipo de vulnerabilidad maliciosa de un sitio web donde se pueden transmitir comandos no autorizados desde el navegador de un usuario en el que la aplicación web confía.

Para abordar esta clase de casos, la aplicación envía un mensaje de advertencia al usuario, solicitando que cambie su escala de datos para que el test estadístico pueda ser calculado.

2.2.5 Fase V. Despliegue

Previo a las revisiones y pruebas internas realizadas a ABAL, donde su funcionamiento operó de manera estable, se decide lanzar la versión “beta abierto” (para el público general). Con esto se busca observar su comportamiento en un ambiente no tan controlado y detectar posibles errores y problemas técnicos, además de recibir retroalimentación del mismo.

Por lo tanto, en esta sección se presenta la documentación técnica de la primera versión de ABAL 1.0Beta, con el objetivo de brindar asistencia a los usuarios que harán uso de esta.

En la figura 2.15 muestra la imagen de la página principal de ABAL. En esta se puede observar un menú con las opciones de Inicio, Ayuda, Contacto y Repositorio. La opción de Inicio permite volver a la página principal cuando se encuentre en otra sección distinta. En la parte de Ayuda, se proporciona al usuario una explicación breve de cada una de las distribuciones presentadas en el contenido de la página principal. En el apartado de Contacto se pone a disposición de los usuarios una cuenta de correo para que expongan sus dudas o sugerencias de la aplicación, y en la opción de Repositorio se podrá consultar un listado de muestras reales que han sido objeto de diversos trabajos en el IMT.



Fuente: Elaborado por el IMT

Figura 2.15 Página principal de ABAL

Continuando en la página principal, se presentan al usuario para sus respectivos cálculos las opciones de distribuciones de probabilidad discretas y continuas. Para el caso de las Discretas se tienen la Binomial, Geométrica número de fallas, Geométrica número de pruebas y Poisson. En el apartado de Continuas se tienen las distribuciones: Uniforme con la prueba Ji-Cuadrada y Kolmogorov, la Normal y Exponencial con las pruebas Ji-Cuadrada, Kolmogorov y Anderson, la Lognormal, Weibull, y Gamma con las pruebas de Ji-Cuadrada y Kolmogorov.

Si se elige la distribución Binomial, la aplicación cambiará de página mostrando ahora la opción de llenado de la información. Como se muestra en la figura 2.16 a manera de ayuda se presentan los parámetros: función de probabilidad, media y varianza correspondientes a esa distribución.

Binomial			
Parámetros	Función de probabilidad	Media	Varianza
$n > 1; 0 < p < 1$	$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$ $k = 0, 1, 2, \dots, n$	np	$np(1-p)$

Fuente: Elaborado por el IMT

Figura 2.16 Página principal de ABAL

Como se muestra en la figura 2.17 se tiene una hoja tipo Excel con 6 columnas (A-F) opcionales para ingresar la muestra de datos. Se observan como mínimo 20 renglones, pero se pueden expandir tanto como se requieran. Para ello se deberá teclear “enter” después de cada número ingresado y automáticamente los renglones de la columna empezarán a incrementarse hacia abajo, o bien, se puede hacer uso de la función Ctrl+c y Ctrl+v de Windows.

Ingrese la muestra de datos						
	A	B	C	D	E	F
1	9					
2	6					
3	6					
4	7					
5	4					
6	8					
7	7					
8	8					
9	4					
10	7					
11	10					
12	9					
13	8					
14	7					
15	9					
16	3					
17	4					
18	6					
19	6					
20	5					

Fuente: Elaborado por el IMT

Figura 2.17 Sección de ingreso de la muestra de datos.

En caso de pegar la muestra con la opción de Windows, solo se mostrarán los primeros veinte renglones. Para desplazarse (arriba, abajo) sobre el contenido y así poder observar el resto de la muestra, se deberá usar la barra situada al costado derecho de la hoja, mejor conocido como *scroll*.

Posteriormente, si se conoce el número de ensayos para esta distribución se podrán ingresar, de lo contrario se podrá dejar el campo de texto como vacío y ABAL procederá a realizar su cálculo. Y para ofrecer un resultado más personalizado, se tiene como opción ingresar el título del histograma, así como, su referencia para el eje x como se muestra en la figura 2.18.

Si conoce el número de ensayos ingreselo:

Número de ensayos:

Especifique las características del histograma a generar:

Título del histograma:

Título en eje x:

Calcular

Fuente: Elaborado por el IMT

Figura 2.18 Sección de ingreso de la muestra de datos.

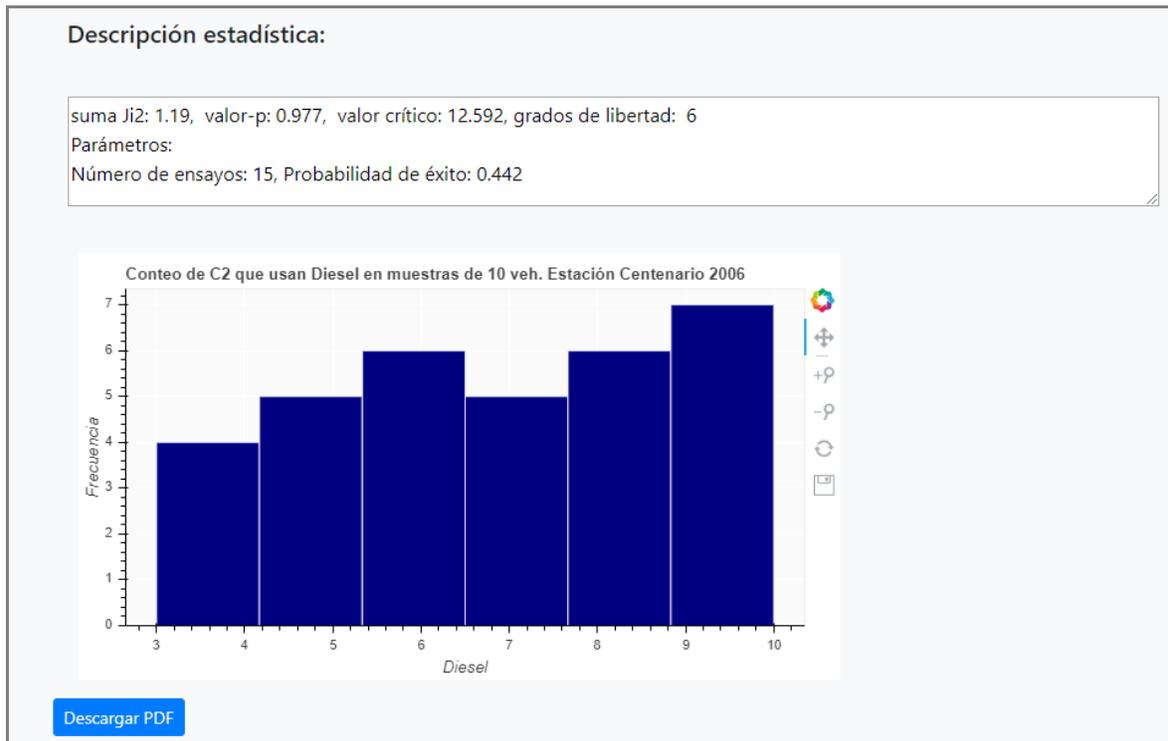
Por último, al dar clic en el botón "Calcular" la aplicación mostrará la descripción estadística de la distribución, como se presenta en la figura 2.18. En este caso, los resultados son: suma Ji^2 , valor-p, valor crítico, grados de libertad, y como parámetros n (número de ensayos) y p (número de pruebas).

Posteriormente se muestra el histograma con los títulos ingresados. Al costado derecho de este, se observa una barra de herramientas con las acciones propias de la librería *Bokeh*. Estas son: "Pan, Zoom In, Zoom Out, Reset, Save". A continuación se describen cada una de estas.

Tabla 2.10 Barra de herramientas de histograma

Ícono	Nombre	Descripción
	<i>Pan</i>	Permite desplazar el trazado arrastrando el mouse.
	<i>Zoom In</i>	Aumenta el zoom de la trama, respetando los valores y rangos mínimos y máximos, evitando el acercamiento y alejamiento más allá de estos.
	<i>Zoom Out</i>	Disminuye el nivel del zoom de la trama. Respeta los valores, así como los rangos mínimos y máximos, evitando el acercamiento y alejamiento más allá de estos.
	<i>Reset</i>	Restaura los rangos de trazado a sus valores originales.
	<i>Save</i>	Muestra un cuadro de dialogo modal que permite al usuario guardar una imagen PNG de la trama.

Fuente: Bokeh. Configuración de herramientas de trazado.



Fuente: Elaborado por el IMT

Figura 2.19 Sección de resultados

Finalmente, se proporciona la opción de descargar un documento en formato PDF (*Portable Document Format*). Para ello, se deberá dar clic en el botón “Descargar PDF” y automáticamente aparecerá en la parte inferior derecha de la página, el archivo identificado con el nombre de la distribución elegida, un ejemplo de esto se

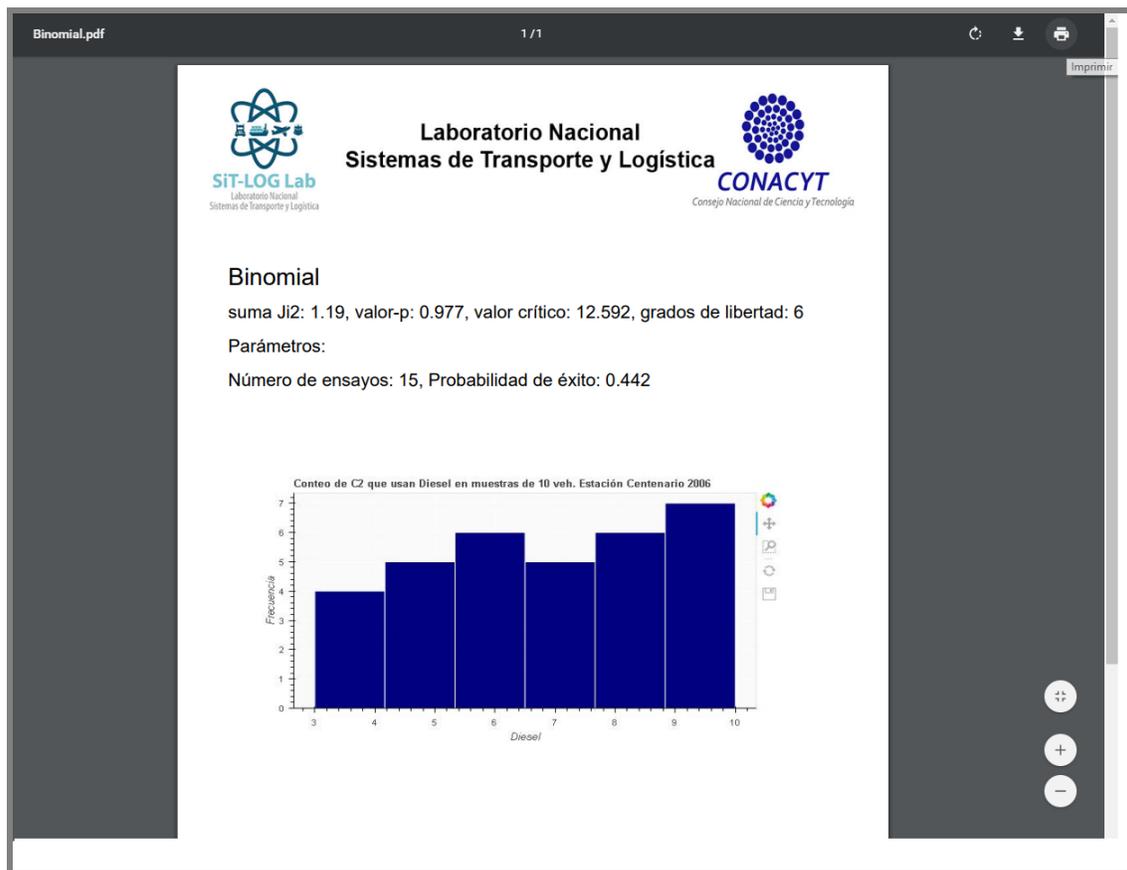
muestra en la figura 2.20. El usuario podrá abrir este documento dando doble clic sobre este, posteriormente aparecerá en una nueva ventana del navegador.



Fuente: Elaborado por el IMT

Figura 2.20 Descarga de documento pdf.

Otra forma de acceder al documento es buscarlo en la carpeta de Descargas (*Downloads*) de Windows, dar doble clic sobre este y automáticamente se abrirá en una nueva ventana del navegador. Desde ahí se podrán realizar las acciones de: imprimir, descargar, acercar, alejar, y ajustar como se muestra en la figura 2.21.



Fuente: Elaborado por el IMT

Figura 2.21 Vista del documento Binomial.pdf

Si se elige cualquier otra distribución, el procedimiento de ingreso de datos es el mismo descrito con anterioridad.

3 Distribuciones de probabilidad en transporte e ingeniería

En las tareas de investigación y de solución de problemas en ingeniería y en transporte aparecen datos de tipo aleatorio que requieren ser ajustados a una distribución de probabilidad que los represente adecuadamente. El primer paso para este ajuste es identificar un modelo probabilístico que pudiera ser usado en los datos de interés.

En este capítulo se hace una breve revisión de las distribuciones de probabilidad de uso más común en transporte e ingeniería, para tener una visión panorámica que le permita al lector elegir entre estas distribuciones y aquellas con posibilidades de representar sus datos.

3.1 Distribuciones discretas

Estas distribuciones sirven para modelar datos aleatorios de tipo discreto, en general medidos con números enteros. Ejemplos de esto son: aforos de pasajeros en una terminal aérea, conteo vehicular en una carretera, número de accidente en un paso ferroviario a nivel, etc.

Cuatro distribuciones que se han usado en trabajos de ingeniería y transporte son:

- a) Binomial
- b) Geométrica
- c) Poisson
- d) Logarítmica

Para cada una de estas distribuciones se tiene una función de probabilidad, que calcula la probabilidad de que la variable aleatoria del modelo (es decir, el dato) tome un valor específico.

La Tabla 3.1 muestra el resumen de estas distribuciones discretas, indicando la función de probabilidad, y las fórmulas de media y varianza.

Tabla 3.1. Resumen de distribuciones discretas (elaboración propia)

Distribución	Función de probabilidad	Media	Varianza
Binomial	$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$ $n > 1; 0 < p < 1; k = 0, 1, 2, \dots, n$	np	$np(1-p)$
Geométrica (núm. de fallas)	$P[X = k] = p(1-p)^k$ $0 < p < 1; k = 0, 1, 2, \dots$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Geométrica (núm. de ensayos)	$P[X = k] = p(1-p)^{k-1}$ $0 < p < 1; k = 1, 2, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson	$P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$ $\lambda > 0; k = 0, 1, 2, \dots$	λ	λ
Logarítmica	$P[X = k] = \frac{A\theta^k}{k}$ $A = \left[\frac{-1}{\ln(1-\theta)} \right]; 0 < \theta < 1; k = 1, 2, \dots,$	$\frac{A\theta}{1-\theta}$	$\frac{A\theta(1-A\theta)}{(1-\theta)^2}$

Fuente: Elaborado por el IMT

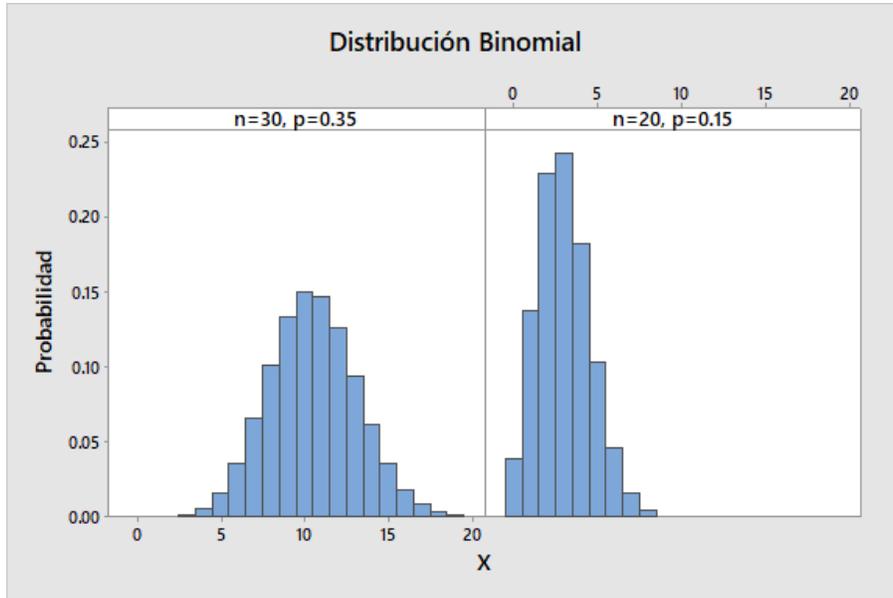
La tabla muestra dos casos para la Geométrica. El modelo probabilístico prototipo que describe la Geométrica supone que hay un evento relevante (llamado “éxito”) con probabilidad p de ocurrir, y que se repite el ensayo hasta que aparece el “éxito”. Las repeticiones donde no aparece “éxito” se llaman “fallas”. La diferencia en las dos versiones está en que:

- a) Se cuenta el número de fallas antes del éxito (primera versión) o
- b) Se cuenta el número de ensayos para llegar al éxito (segunda versión).

El modelador de los datos aleatorios debe decidir cuál versión de estas distribuciones es conveniente para su problema.

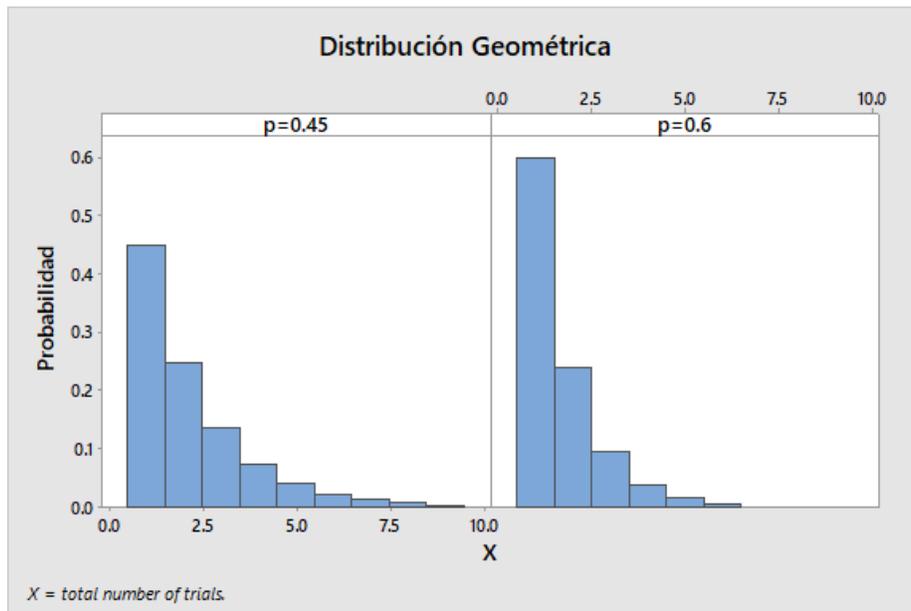
Por ejemplo, si sus datos no contienen ningún cero y le interesa modelar como Geométrica, debe probar con la segunda versión, donde el dato representa el número de ensayos necesario para llegar al éxito, cuyo valor mínimo es 1.

Las siguientes gráficas muestran ejemplos de algunas de estas distribuciones.



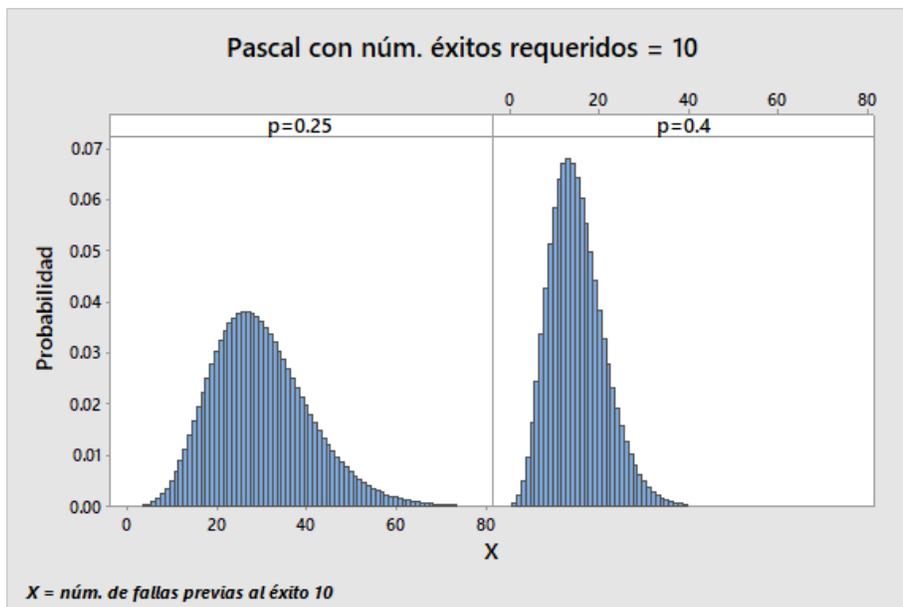
Fuente: Elaborado por el IMT

Figura 3.1 Ejemplos de distribución Binomial (elaboración propia)



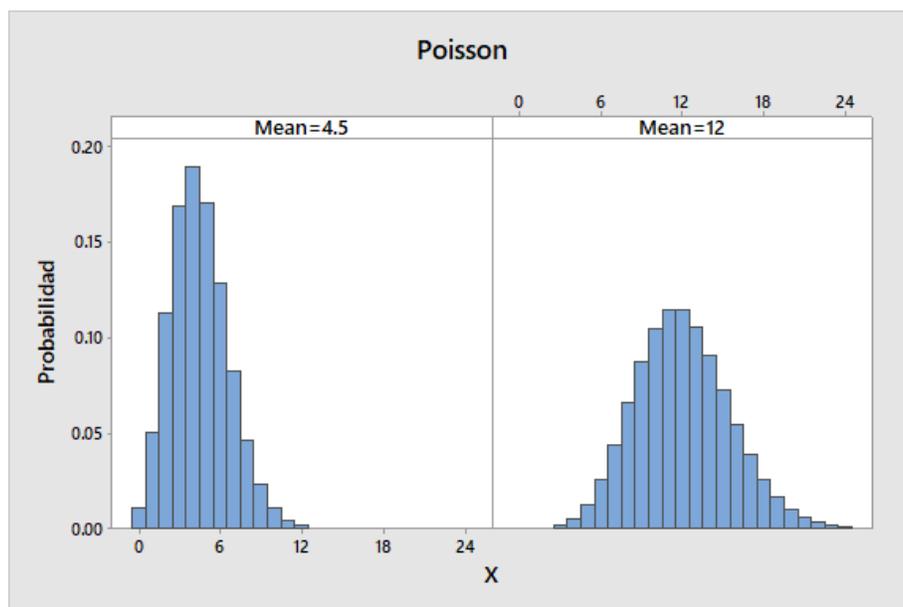
Fuente: Elaborado por el IMT

Figura 3.2 Ejemplos de distribución Geométrica



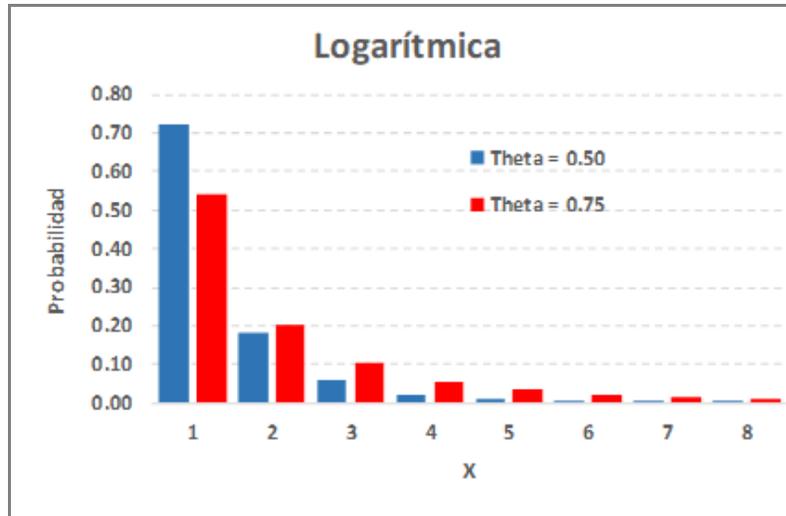
Fuente: Elaborado por el IMT

Figura 3.3 Ejemplos de distribución Pascal



Fuente: Elaborado por el IMT

Figura 3.4 Ejemplos de distribución Poisson



Fuente: Elaborado por el IMT

Figura 3.5 Ejemplos de distribución Logarítmica

Las gráficas previas muestran el comportamiento de las distribuciones discretas al variar sus parámetros. Las formas de los histogramas de los datos para un proyecto, pueden sugerir analogías con algunas distribuciones para identificar modelos probables.

3.2 Distribuciones continuas

Estas distribuciones pueden modelar datos aleatorios de tipo continuo, que en la práctica son números con punto decimal y una expansión decimal finita de la parte fraccionaria. Ejemplos de esto son: tonelajes movidos en camiones articulados, tiempos de travesía en rutas carretera, demoras en la llegada de vuelos en un aeropuerto, etc.

Siete distribuciones comunes en trabajos de ingeniería y transporte son:

- a) Normal
- b) Exponencial
- c) Gamma
- d) Lognormal
- e) Weibull
- f) Logística
- g) Gumbel

Para cada una de estas distribuciones se tiene una función de densidad probabilidad, que permite calcular la probabilidad de que la variable aleatoria del modelo (es decir, el dato) tome valores en un intervalo específico.

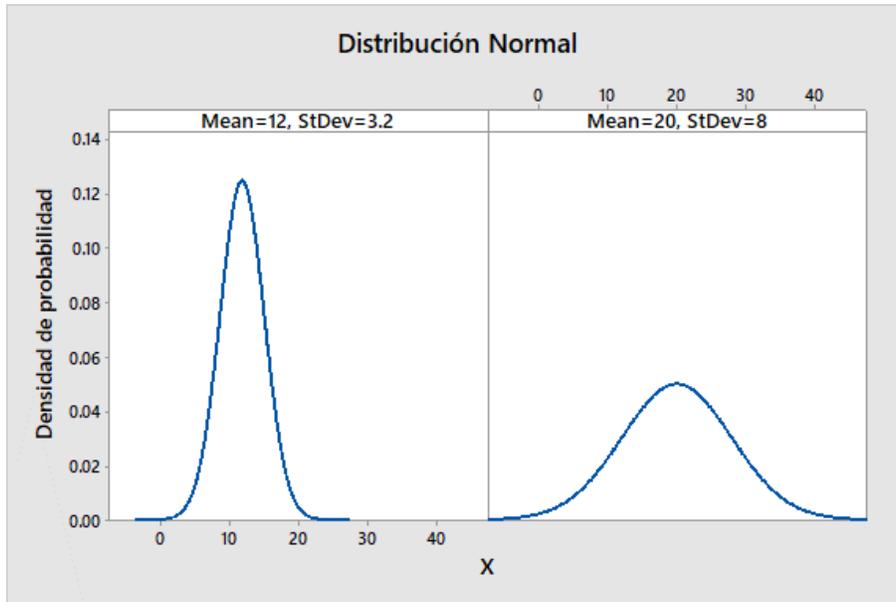
La tabla 3.2 muestra el resumen de estas distribuciones, con sus parámetros y fórmulas de media y varianza.

Tabla 3.2 Resumen de distribuciones continuas

Distribución	Función de densidad de probabilidad	Media	Varianza
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $-\infty < x < \infty$ $-\infty < \mu < \infty; \sigma > 0$	μ	σ^2
Exponencial	$f(x) = \lambda e^{-\lambda x}; x \geq 0$ $\lambda > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma	$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)}$ $\alpha: \text{forma}, \beta: \text{escala}$ $\alpha, \beta > 0; x \geq 0;$	$\alpha\beta$	$\alpha\beta^2$
Lognormal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$ $x, \sigma > 0; -\infty < \mu < \infty$	$e^{\left(\mu + \frac{\sigma^2}{2}\right)}$	$e^{2\sigma^2+2\mu} - e^{\sigma^2+2\mu}$
Weibull	$f(x) = \frac{\lambda}{k} \left(\frac{x}{k}\right)^{\lambda-1} e^{-(x/k)^\lambda}$ $\lambda: \text{forma}, k: \text{escala};$ $\lambda, k > 0; x > 0$	$k\Gamma\left(1 + \frac{1}{\lambda}\right)$	$k^2\Gamma\left(1 + \frac{2}{\lambda}\right) - \left[\Gamma\left(1 + \frac{1}{\lambda}\right)\right]^2$
Logística	$f(x) = \frac{e^{-\left(\frac{x-a}{b}\right)}}{b \left[1 + e^{-\left(\frac{x-a}{b}\right)}\right]^2}$ $-\infty < x < \infty$ $-\infty < a < \infty; b > 0$	a	$\frac{b^2\pi^2}{3}$
Gumbel	$f(x) = \frac{1}{b} e^{-\left[\frac{x-a}{b} + e^{-\left(\frac{x-a}{b}\right)}\right]}$ $-\infty < x < \infty$ $-\infty < a < \infty; b > 0$	$a + \gamma b$ $\gamma \approx 0.5772$	$\frac{b^2\pi^2}{6}$

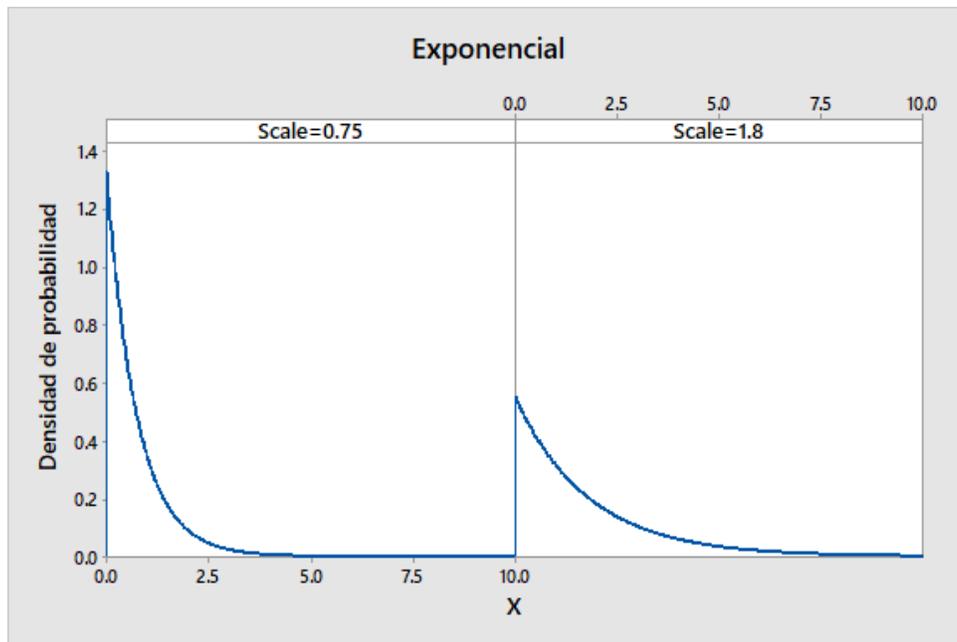
Fuente: Elaborado por el IMT

Las siguientes gráficas muestran ejemplos de densidades para algunas de estas distribuciones.



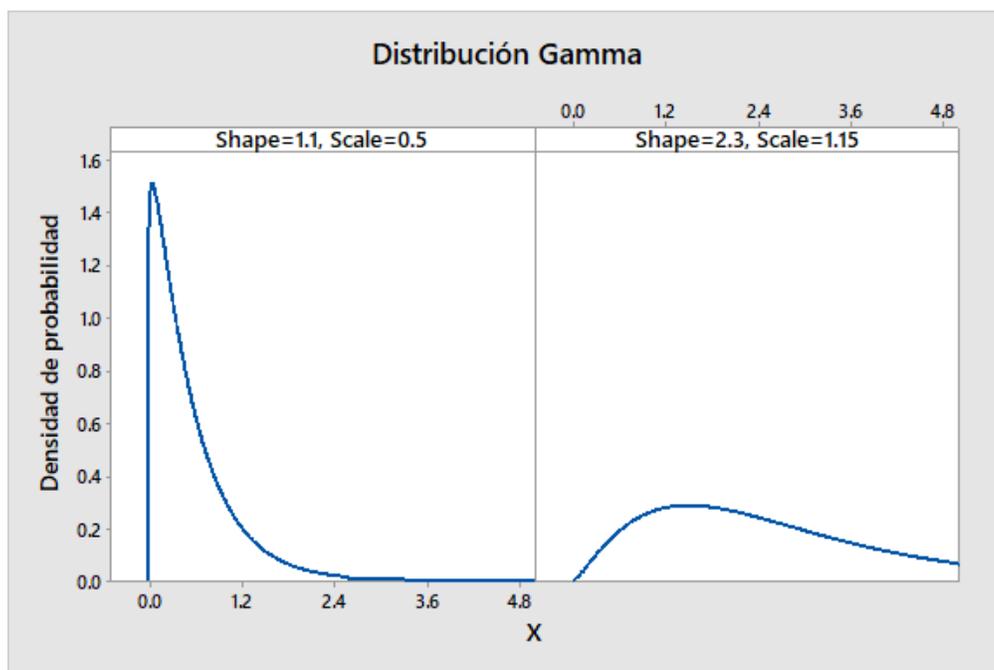
Fuente: Elaborado por el IMT

Figura 3.6 Ejemplos de distribución Normal



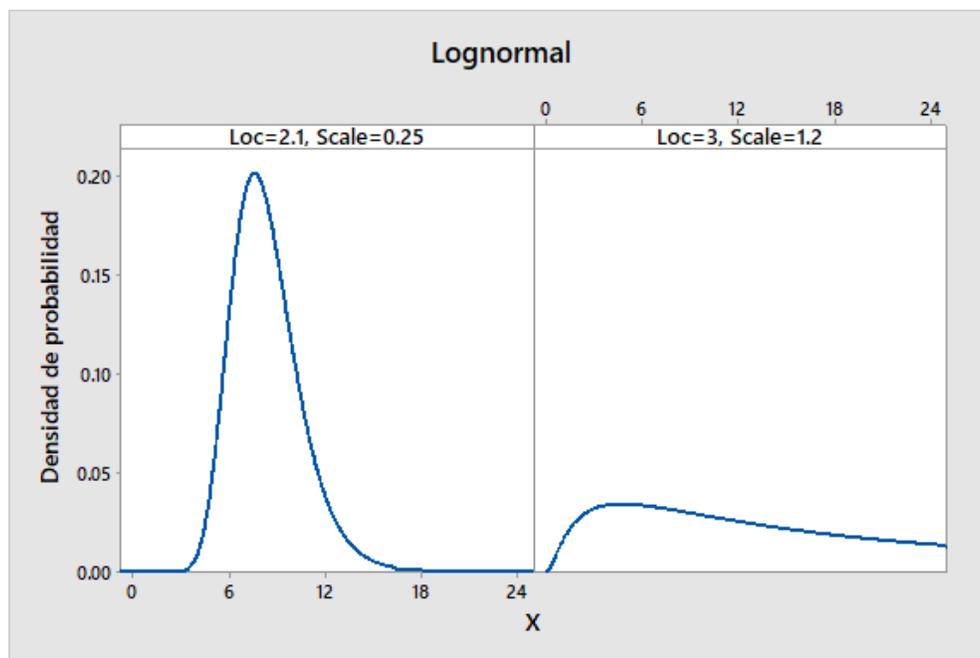
Fuente: Elaborado por el IMT

Figura 3.7 Ejemplos de distribución Exponencial



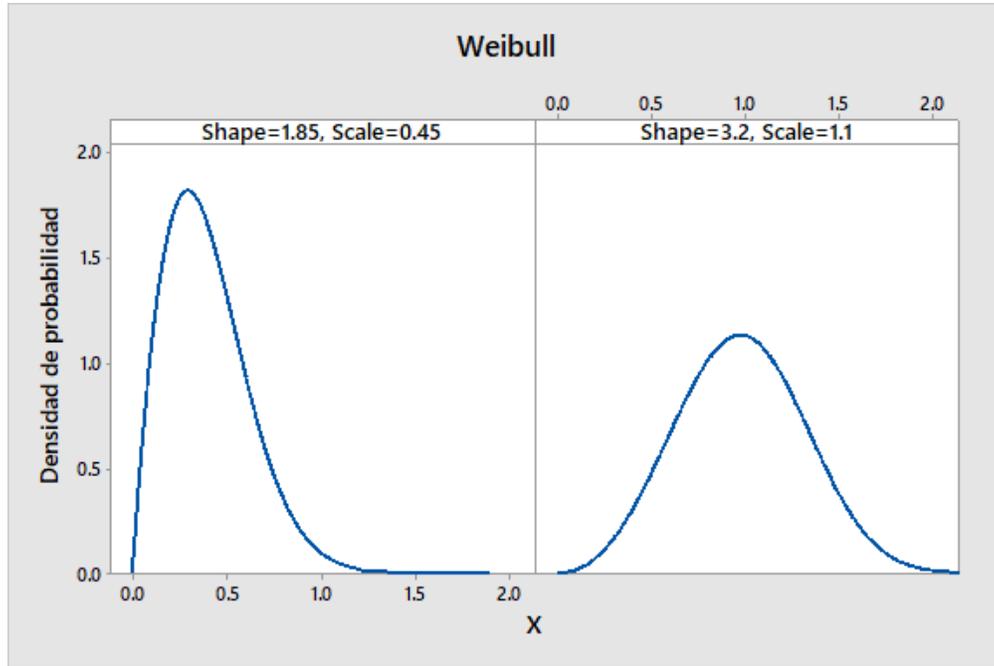
Fuente: Elaborado por el IMT

Figura 3.8 Ejemplos de distribución Gamma



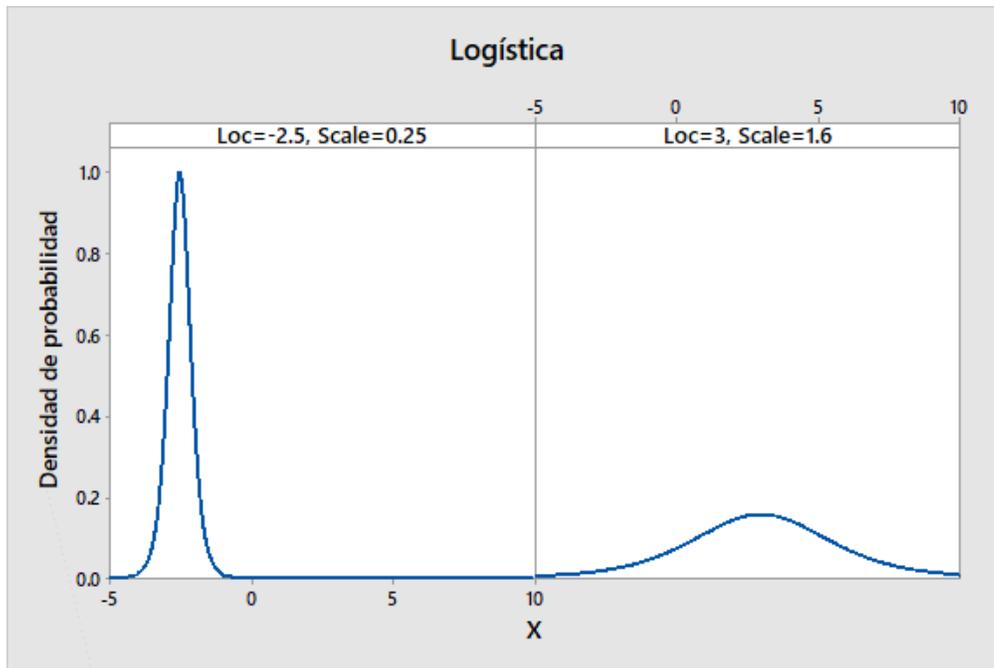
Fuente: Elaborado por el IMT

Figura 3.9 Ejemplos de distribución Lognormal



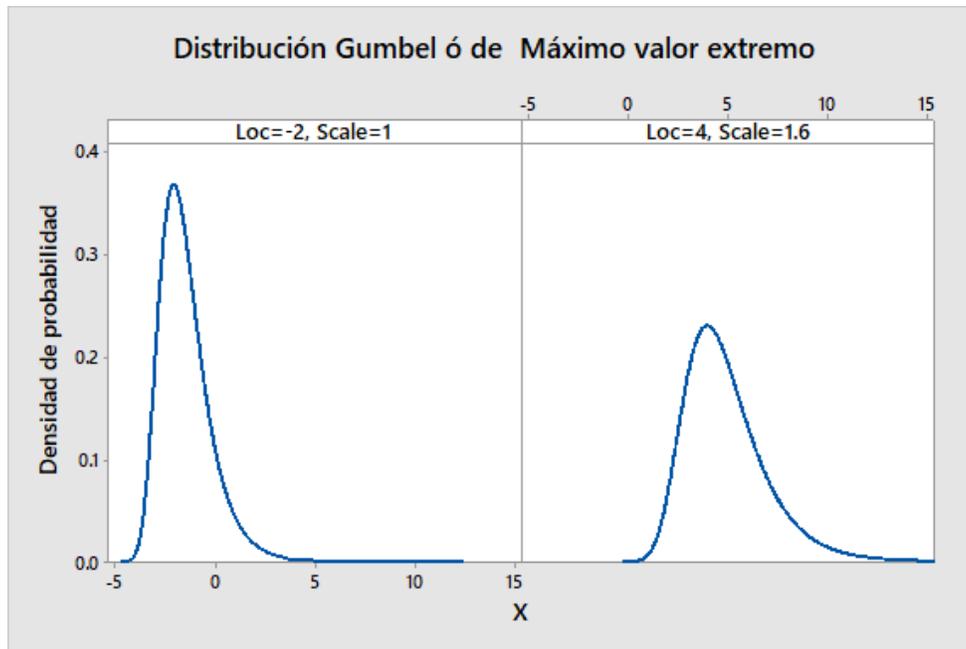
Fuente: Elaborado por el IMT

Figura 3.10 Ejemplos de distribución Weibull



Fuente: Elaborado por el IMT

Figura 3.11 Ejemplos de distribución Logística



Fuente: Elaborado por el IMT

Figura 3.12 Ejemplos de distribución Gumbel

4 Técnicas de Bondad de Ajuste

Las técnicas de bondad de ajuste son pruebas de hipótesis estadísticas que dan un criterio para aceptar o rechazar el modelo probabilístico que se haya propuesto para los datos de interés. La elección del modelo probabilístico que pudiera ajustarse a un conjunto de datos dado, suele en la práctica tomar en cuenta los siguientes aspectos.

1. La forma del histograma de los datos puede sugerir alguna distribución. Por ejemplo, si el histograma es de datos continuos y luce acampanado con simetría central, las distribuciones Normal o Logística podrían elegirse; si los datos fueran discretos, la Binomial o la Poisson pudieran funcionar.
2. La información histórica de distribuciones usadas en ciertas áreas puede ser útil. Por ejemplo, en confiabilidad y riesgo en sistemas/equipos, las distribuciones Exponencial y Weibull se usan mucho; en Hidrología (niveles de lluvia, caudales de ríos, etc.) la Lognormal y la Gamma se han usado con éxito. En conteos de accidentes y aforos vehiculares, las distribuciones Poisson y Binomial Negativa son de uso común.
3. Algunas razones teóricas con ciertas distribuciones. Por ejemplo, se ha demostrado que en el conteo de eventos con distribución Poisson, los tiempos que transcurren entre eventos se distribuyen como Exponencial; también, al muestrear repetidamente una medición continua, el Teorema del Límite Central asegura que los promedios siguen una distribución Normal.

Una vez elegida una distribución de probabilidad para ajustar los datos, se deben resolver dos tareas:

- a) Estimar los parámetros de la distribución, usando los datos, por ejemplo, si la propuesta es una Poisson, hay que estimar el parámetro λ ; si se trata de una Lognormal, hay que estimar los valores de μ y σ correspondientes, etc.
- b) Realizar la prueba de bondad de ajuste para los datos, suponiendo que provienen de la distribución propuesta, y decidir si se acepta o se rechaza.

La estimación de los parámetros se hace en general con dos métodos, el de momentos y el de máxima verosimilitud. Los detalles de su cálculo se pueden encontrar en Moreno, E. y Nieves, V. (2019). En la tabla 4.1 se muestran las fórmulas para estimar los parámetros de diversas distribuciones, utilizando la muestra de datos que se desea ajustar.

Para aplicar estas fórmulas, deben calcularse primero los momentos muestrales usando los datos que se desean ajustar. Si se tiene una muestra de datos $X_1, X_2, X_3, \dots, X_n$, los momentos muestrales de orden 1 y 2 (m'_1 y m'_2) calculados a partir de los datos $X_1, X_2, X_3, \dots, X_n$ son como sigue:

$$m'_1 = \frac{1}{n} \sum_{j=1}^n X_j \quad m'_2 = \frac{1}{n} \sum_{j=1}^n X_j^2$$

Tabla 4.1 Estimadores de parámetros de algunas distribuciones

Distribución	Parámetros	Estimadores
Binomial	$n > 1$ $0 < p < 1$	$n = \frac{m'_1{}^2}{m'_1 + m'_1{}^2 - m'_2}$ $p = \frac{m'_1 + m'_1{}^2 - m'_2}{m'_1}$
Geométrica (No. de fallas)	$0 < p < 1$	$p = \frac{1}{1 + m'_1}$
Geométrica (No. ensayos)	$0 < p < 1$	$p = \frac{1}{m'_1}$
Poisson	$\lambda > 0$	$\lambda = m'_1$
Normal	$-\infty < \mu < \infty$ $\sigma > 0$	$\mu = m'_1; \sigma = \sqrt{m'_2 - m'_1{}^2}$
Exponencial	$\lambda > 0$	$\lambda = \frac{1}{m'_1}$
Gamma	$\alpha, \beta > 0$	$\alpha = \frac{m'_1{}^2}{m'_2 - m'_1{}^2}; \beta = \frac{m'_2 - m'_1{}^2}{m'_1}$
Lognormal	$x, \sigma > 0$ $-\infty < \mu < \infty$	De los datos X_i se calcula $Y_i = \ln X_i$ $\mu = E(Y) = m'_{1Y}$ $\sigma = \sqrt{E(Y^2) - E(Y)^2} = \sqrt{m'_{2Y} - m'_{1Y}{}^2}$

Fuente: Elaborado por el IMT

Ya que se han estimado los parámetros de la distribución que se propone para los datos, se realiza alguna prueba de bondad de ajuste, para tener un criterio estadístico para aceptar o rechazar la propuesta. En la sección siguiente se revisan tres técnicas de bondad de ajuste que se usan comúnmente para esto:

- a. Prueba Ji-Cuadrada (χ^2)
- b. Prueba de Kolmogorov-Smirnov
- c. Prueba de Anderson-Darling

4.1 La prueba Ji-Cuadrada

Esta técnica compara la frecuencia observada en los datos contra la frecuencia esperada conforme a la distribución de probabilidad propuesta para los datos. Las diferencias entre frecuencia observada y la esperada con la distribución teórica se elevan al cuadrado y se suman, y la suma se somete a una prueba de hipótesis con una distribución Ji-Cuadrada, para decidir si la distribución propuesta es adecuada o no, usualmente con un nivel de significación del 5%.

Esta prueba funciona con datos discretos y con datos continuos. Se recomienda tener al menos $N=30$ de tamaño de muestra. Con muestras más chicas, la prueba puede funcionar, pero si hay pocos datos, los resultados son menos confiables.

Con una muestra de N datos, la prueba se desarrolla con los siguientes pasos.

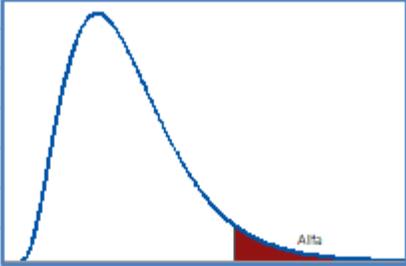
1. Los N datos se ponen en una tabla de frecuencias con $m \approx \sqrt{N}$ intervalos y se obtiene la FRECUENCIA OBSERVADA FO_j en cada intervalo j
2. Para la distribución de probabilidad propuesta, se define $k = \text{núm. de parámetros necesarios para determinar la distribución}$ y se estiman sus parámetros a partir de la muestra.
Con esta distribución se calcula la FRECUENCIA ESPERADA FE_j en cada intervalo j como: $FE_j = P[\text{el dato esté en el intervalo } j] \times N$
3. Se calcula el estimador:

$$C = \sum_{j=1}^m \frac{(FE_j - FO_j)^2}{FE_j}$$

4. Si C es menor o igual al valor crítico de una χ^2 con $m - k - 1$ grados de libertad, a un nivel de significación del $\alpha\%$, no se puede rechazar la hipótesis de que los datos corresponden a la distribución propuesta.

La prueba Ji-Cuadrada puede usarse tanto para distribuciones discretas como continuas. La tabla 4.2 enseguida muestra valores críticos para la prueba Ji-Cuadrada.

Tabla 4.2 Valores críticos prueba Ji cuadrada.

									
		Alfa ---->							
Gr. Lib		0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025
1		0.000	0.001	0.004	0.016	0.455	2.706	3.841	5.024
2		0.020	0.051	0.103	0.211	1.386	4.605	5.991	7.378
3		0.115	0.216	0.352	0.584	2.366	6.251	7.815	9.348
4		0.297	0.484	0.711	1.064	3.357	7.779	9.488	11.143
5		0.554	0.831	1.145	1.610	4.351	9.236	11.070	12.833
6		0.872	1.237	1.635	2.204	5.348	10.645	12.592	14.449
7		1.239	1.690	2.167	2.833	6.346	12.017	14.067	16.013
8		1.646	2.180	2.733	3.490	7.344	13.362	15.507	17.535
9		2.088	2.700	3.325	4.168	8.343	14.684	16.919	19.023
10		2.558	3.247	3.940	4.865	9.342	15.987	18.307	20.483
11		3.053	3.816	4.575	5.578	10.341	17.275	19.675	21.920
12		3.571	4.404	5.226	6.304	11.340	18.549	21.026	23.337
13		4.107	5.009	5.892	7.042	12.340	19.812	22.362	24.736
14		4.660	5.629	6.571	7.790	13.339	21.064	23.685	26.119
15		5.229	6.262	7.261	8.547	14.339	22.307	24.996	27.488
16		5.812	6.908	7.962	9.312	15.338	23.542	26.296	28.845
17		6.408	7.564	8.672	10.085	16.338	24.769	27.587	30.191
18		7.015	8.231	9.390	10.865	17.338	25.989	28.869	31.526
19		7.633	8.907	10.117	11.651	18.338	27.204	30.144	32.852
20		8.260	9.591	10.851	12.443	19.337	28.412	31.410	34.170

Fuente: Elaborado por el IMT

4.2 La prueba de Kolmogorov-Smirnov

Esta técnica compara la Función de Distribución Acumulada (FDA) empírica (observada) obtenida de los datos, con la FDA teórica de acuerdo a la distribución de probabilidad propuesta.

Una limitación de esta prueba es que solo debe usarse para distribuciones continuas. Se recomienda tener muestras de datos de al menos $N=30$; con muestras más pequeñas, la prueba puede funcionar, pero con pocos datos, los resultados son menos confiables.

Partiendo de una muestra de N datos, y con la distribución propuesta para la prueba los pasos son como sigue:

- 1) Los N datos se escriben *en orden creciente* en una tabla: X_1, X_2, \dots, X_N y se estiman los parámetros de la distribución propuesta para los datos.
- 2) Se calcula la *probabilidad acumulada observada* PAO_j de cada dato X_j
- 3) Con la distribución teórica propuesta se calcula la *probabilidad acumulada esperada* PAE_j para cada dato X_j .
- 4) Se calcula el valor absoluto $|PAO_j - PAE_j|$ para cada dato X_j y se busca la máxima diferencia hallada $MD = \text{Max}\{|PAO_j - PAE_j|, j = 1, 2, \dots, N\}$ en la tabla de datos.
- 5) El estimador MD se compara con el valor crítico de la *Tabla K-S* con N datos y un nivel de significación del $\alpha\%$.
Si MD es menor o igual que este valor crítico, entonces no se puede rechazar la hipótesis de que los datos corresponden a la distribución propuesta.

Los valores críticos para varios niveles de significación de la prueba Kolmogorov-Smirnov se muestran en la tabla 4.3 enseguida.

También pueden encontrarse en sitios de temas estadísticos en Internet, como es el caso de *Real Statistics Using Excel* en: <http://www.real-statistics.com/statistics-tables/kolmogorov-smirnov-table/>

Tabla 4.3 Valores críticos. Prueba Kolmogorov-Smirnov. (Frías, M.P., 2018)

N	NIVEL DE SIGNIFICACIÓN α							
	20.0%	10.0%	5.0%	2.0%	1.0%	0.5%	0.2%	0.1%
1	0.90000	0.95000	0.97500	0.99000	0.99500	0.99750	0.99900	0.99950
2	0.68337	0.77639	0.84189	0.90000	0.92929	0.95000	0.96838	0.97764
3	0.56481	0.63604	0.70760	0.78456	0.82900	0.86428	0.90000	0.92065
4	0.49265	0.56522	0.62394	0.68887	0.73424	0.77639	0.82217	0.85047
5	0.44698	0.50945	0.56328	0.62718	0.66853	0.70543	0.75000	0.78137
6	0.41037	0.46799	0.51926	0.57741	0.61661	0.65287	0.69571	0.72479
7	0.38148	0.43607	0.48342	0.53844	0.57581	0.60975	0.65071	0.67930
8	0.35831	0.40962	0.45427	0.50654	0.54179	0.57429	0.61368	0.64098
9	0.33910	0.38746	0.43001	0.47960	0.51332	0.54443	0.58210	0.60846
10	0.32260	0.36866	0.40925	0.45562	0.48893	0.51872	0.55500	0.58042
11	0.30829	0.35242	0.39122	0.43670	0.46770	0.49539	0.53135	0.55588
12	0.29577	0.33815	0.37543	0.41918	0.44905	0.47672	0.51047	0.53422
13	0.28470	0.32549	0.36143	0.40362	0.43247	0.45921	0.49189	0.51490
14	0.27481	0.31417	0.34890	0.38970	0.41762	0.44352	0.47520	0.49753
15	0.26589	0.30397	0.33750	0.37713	0.40420	0.42934	0.45611	0.48182
16	0.25778	0.29472	0.32733	0.36571	0.39201	0.41644	0.44637	0.46750
17	0.25039	0.28627	0.31796	0.35528	0.38086	0.40464	0.43380	0.45540
18	0.24360	0.27851	0.30936	0.34569	0.37062	0.39380	0.42224	0.44234
19	0.23735	0.27136	0.30143	0.33685	0.36117	0.38379	0.41156	0.43119
20	0.23156	0.26473	0.29408	0.32866	0.35241	0.37451	0.40165	0.42085
21	0.22517	0.25858	0.28724	0.32104	0.34426	0.36588	0.39243	0.41122
22	0.22115	0.25283	0.28087	0.31394	0.33666	0.35782	0.38382	0.40223
23	0.21646	0.24746	0.27491	0.30728	0.32954	0.35027	0.37575	0.39380
24	0.21205	0.24242	0.26931	0.30104	0.32286	0.34318	0.36787	0.38588
25	0.20790	0.23768	0.26404	0.29518	0.31657	0.33651	0.36104	0.37743
26	0.20399	0.23320	0.25908	0.28962	0.30963	0.33022	0.35431	0.37139
27	0.20030	0.22898	0.25438	0.28438	0.30502	0.32425	0.34794	0.36473
28	0.19680	0.22497	0.24993	0.27942	0.29971	0.31862	0.34190	0.35842
29	0.19348	0.22117	0.24571	0.27471	0.29466	0.31327	0.33617	0.35242
30	0.19032	0.21756	0.24170	0.27023	0.28986	0.30818	0.33072	0.34672
31	0.18732	0.21412	0.23788	0.26596	0.28529	0.30333	0.32553	0.34129
32	0.18445	0.21085	0.23424	0.26189	0.28094	0.29870	0.32058	0.33611
33	0.18171	0.20771	0.23076	0.25801	0.27577	0.29428	0.31584	0.33115
34	0.17909	0.21472	0.22743	0.25429	0.27271	0.29005	0.31131	0.32641
35	0.17659	0.20185	0.22425	0.25073	0.26897	0.28600	0.30597	0.32187
36	0.17418	0.19910	0.22119	0.24732	0.26532	0.28211	0.30281	0.31751
37	0.17188	0.19646	0.21826	0.24404	0.26180	0.27838	0.29882	0.31333
38	0.16966	0.19392	0.21544	0.24089	0.25843	0.27483	0.29498	0.30931
39	0.16753	0.19148	0.21273	0.23785	0.25518	0.27135	0.29125	0.30544
40	0.16547	0.18913	0.21012	0.23494	0.25205	0.26803	0.28772	0.30171
41	0.16349	0.18687	0.20760	0.23213	0.24904	0.26482	0.28429	0.29811
42	0.16158	0.18468	0.20517	0.22941	0.24613	0.26173	0.28097	0.29465
43	0.15974	0.18257	0.20283	0.22679	0.24332	0.25875	0.27778	0.29130
44	0.15795	0.18051	0.20056	0.22426	0.24060	0.25587	0.27468	0.28806
45	0.15623	0.17856	0.19837	0.22181	0.23798	0.25308	0.27169	0.28493
46	0.15457	0.17665	0.19625	0.21944	0.23544	0.25038	0.26880	0.28190
47	0.15295	0.17481	0.19420	0.21715	0.23298	0.24776	0.26600	0.27896
48	0.15139	0.17301	0.19221	0.21493	0.23059	0.24523	0.26328	0.27611
49	0.14987	0.17128	0.19028	0.21281	0.22832	0.24281	0.26069	0.27339
50	0.14840	0.16959	0.18841	0.21068	0.22604	0.24039	0.25809	0.27067
n > 50	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$	$\frac{1.73}{\sqrt{n}}$	$\frac{1.85}{\sqrt{n}}$	$\frac{1.95}{\sqrt{n}}$

4.3 La prueba de Anderson-Darling

Esta prueba mejora el desempeño de la prueba K-S, pues detecta mejor las diferencias entre las distribuciones acumuladas empírica y teórica en las colas de las gráficas. Se recomienda usar solamente para distribuciones continuas.

Una particularidad de la prueba de Anderson-Darling es que sus valores críticos no se encuentran en una sola tabla, como pasa con la Ji-Cuadrada o la Kolmogorov-Smirnov, sino que hay una tabla específica para cada distribución.

La prueba Anderson-Darling (AD) es válida solo para distribuciones continuas. Los pasos de ejecución son:

- 1) Con la muestra de N datos se estiman los parámetros de la distribución propuesta, y se realizan dos ordenamientos de datos:
- 2) Los datos de *menor a mayor*: $Y_j, j = 1, 2, \dots, N; Y_k \leq Y_{k+1}$
- 3) Los datos de *mayor a menor* $Y_{N-j+1}, j = 1, 2, \dots, N; Y_k \geq Y_{k+1}$.
- 6) Se calcula la *probabilidad acumulada esperada* PAE_j para cada dato Y_j . y la *probabilidad acumulada esperada* PAE_{N+1-j} para cada dato Y_{N+1-j} . para cada para cada valor Y_{n+j-1} con base en la distribución de probabilidad teórica propuesta para la prueba.
- 4) El estadístico de la prueba se calcula como:

$$A_N = - \left[N + \frac{1}{N} \sum_{j=1}^N (2j-1) [LnPAE_j + Ln(1 - PAE_{N-j+1})] \right]$$

- 5) Se ajusta el estadístico de prueba conforme a la distribución de probabilidad teórica propuesta, según la tabla de valores críticos Anderson-Darling, y se determina el valor crítico de la prueba.
- 6) Si el estadístico de prueba es menor que el valor crítico del punto anterior, no se puede rechazar la hipótesis nula; en caso contrario, se rechaza.

Los valores críticos de algunas distribuciones comunes para la prueba AD se muestran en la Tabla 4.4 enseguida.

Tabla 4.4 Valores críticos. Prueba Anderson-Darling.

Distribución	Estadístico ajustado	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
Parámetros conocidos, $N \geq 5$	A_N^2	1.933	2.492	3.070	3.853
Parámetros estimados a partir de los datos					
Normal y Lognormal	$A_N^2 \left(1 + \frac{3}{4N} + \frac{9}{4N^2} \right)$	0.631	0.752	0.873	1.035
Exponencial	$A_N^2 \left(1 + \frac{3}{10N} \right)$	1.062	1.321	1.591	1.959
Weibull	$A_n^2 \left(1 + \frac{1}{5\sqrt{n}} \right)$	0.637	0.757	0.877	1.038

Fuente: Stephens, 1979; Real Statistics, 2018

5 Ejemplos de uso de la aplicación con datos reales

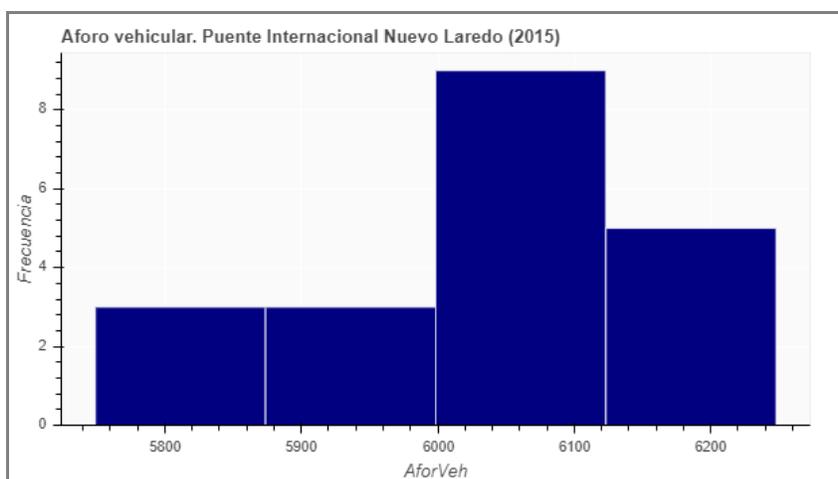
Los ejemplos numéricos enseguida se elaboraron utilizando datos reales de diversas fuentes relacionadas a problemáticas del transporte. Cada caso muestra el desarrollo siguiendo los pasos de las respectivas pruebas de bondad de ajuste que se describieron en el capítulo 4; así mismo, las comparaciones de los valores críticos utilizan las tablas proporcionadas en ese mismo capítulo.

Ejemplo 5.1 La tabla 5.1 son datos de aforos vehiculares en el Puente Internacional Nuevo Laredo del año 2015; la figura 5.1 muestra el histograma correspondiente.

Tabla 5.1 Aforos en Puente Internacional Nuevo Laredo (2015)

Aforo vehicular			
6248	6101	5977	6173
6112	6097	6,021	6047
6069	6064	6,070	6145
5986	5803	5994	5749
5756	6216	6167	6076

Fuente: Elaborado por el IMT



Fuente: Elaborado por el IMT

**Figura 5.1 Aforo vehicular. Puente Internacional Nuevo Laredo (2015).
(elaboración propia con ABAL)**

La forma del histograma y el carácter entero (discreto) de los aforos sugiere una distribución Poisson para el ajuste. La media de valores de la tabla : $\mu = 6043.55$ es la estimación para el parámetro λ de la Poisson. La prueba Ji-Cuadrada para estos datos se muestra en la tabla 5.2.

Los valores de frecuencia observada (Frec.Obs.) se obtienen de contar los datos contenidos en los intervalos que se indican. Son un total de 20 datos, por lo que la frecuencia esperada (Frec.Esp.) se obtiene calculando la probabilidad Poisson con parámetro $\lambda = 6043.55$ de que el dato esté en el intervalo indicado, multiplicada por la cantidad de datos: $Frec. Esp. = P[X \text{ esté en el intervalo}] \times \text{Núm. de datos}$

Tabla 5.2 Detalle de la prueba Ji-Cuadrada. Datos de aforos.

Intervalo	De	a:	Frec. Obs.	Frec. Esp.	JiCuad
1	0	5986	5	4.6364	0.0285
2	5987	6069	5	7.9927	1.1205
3	6070	6112	5	3.6222	0.5241
4	6113	6248	5	3.6617	0.4891
5	6249	y más	0	0.0870	0.0870
				Suma	2.2492

Fuente: Elaborado por el IMT

Con 5 intervalos y un parámetro estimado (λ), la Ji cuadrada es de $5 - 1 - 1 = 3$ grados de libertad. El resultado de la prueba en la tabla es 2.2492, con valor-p = 0.5223, calculado con la función inversa de Ji cuadrada con 3 grados de libertad. El resultado indica que los datos ajustan bien a una distribución Poisson (6043.55) con un nivel de confianza del 95%.

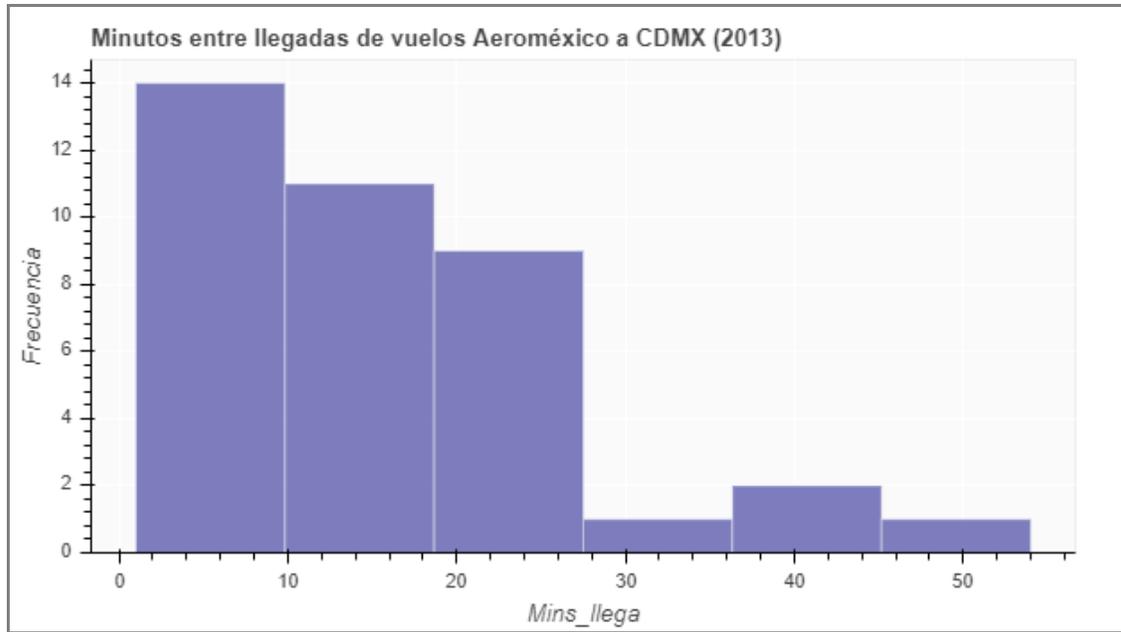
Ejemplo 5.2

Los datos de la tabla 5.3 son minutos entre llegadas de Aeroméxico al aeropuerto de la Ciudad de México, entre las 06:00 y las 15:59 colectados en 2013. El histograma correspondiente se muestra en la figura 5.2.

Tabla 5.3 Minutos entre llegadas de Aeroméxico a CDMX (2013).

Minutos entre llegadas de Aeroméxico (2013)				
9.0	2.0	25.0	8.0	12.0
1.0	19.0	1.0	21.0	9.0
43.0	21.0	16.0	16.0	17.0
36.0	6.0	22.0	14.0	20.0
17.0	12.0	8.0	54.0	18.0
19.0	1.0	5.0	19.0	13.0
39.0	11.0	1.0	22.0	
8.0	3.0	9.0	16.0	

Fuente: Elaborado por el IMT



Fuente: Elaborado por el IMT

**Figura 5.2 Minutos entre llegadas de Aeroméxico a CDMX (2013).
(elaboración propia con ABAL)**

La forma del histograma sugiere la distribución exponencial; su parámetro λ se calcula como el recíproco de la media muestral, resultando así $\lambda = 0.641$.

En la tabla 5.4 se organiza la información para la prueba K-S. La primera columna muestra el dato obtenido; la segunda columna muestra la frecuencia con que aparece en la muestra; la columna 3 muestra esta frecuencia acumulada, es decir, el número de datos menores o iguales al valor referido en la columna de minutos. Las columnas de la función de distribución acumulada muestran los valores en porcentaje acumulado tanto para lo observado en la muestra como para el correspondiente porcentaje que se esperaría bajo el supuesto de que los datos obedecen una ley de probabilidad Expo (0.641).

La prueba K-S calcula el valor absoluto de la diferencia entre la función de distribución acumulada que se observa de los datos y la correspondiente función de distribución esperada, identificando el valor máximo de estas diferencias.

Tabla 5.4 Detalle de la prueba K-S. Datos de llegadas de Aeroméxico

			Función de	Distrib. Acum	Diferencia
Minutos	Frecuencia	Frec. Acum.	Observada	Esperada	ABS(Obs-Esp)
1.0	4	4	0.1053	0.0621	0.0432
2.0	1	5	0.1316	0.1203	0.0113
3.0	1	6	0.1579	0.1749	0.0170
5.0	1	7	0.1842	0.2741	0.0899
6.0	1	8	0.2105	0.3192	0.1087
8.0	3	11	0.2895	0.4011	0.1116
9.0	3	14	0.3684	0.4383	0.0698
11.0	1	15	0.3947	0.5058	0.1111
12.0	2	17	0.4474	0.5365	0.0891
13.0	1	18	0.4737	0.5653	0.0916
14.0	1	19	0.5000	0.5923	0.0923
16.0	3	22	0.5789	0.6413	0.0624
17.0	2	24	0.6316	0.6636	0.0320
18.0	1	25	0.6579	0.6845	0.0266
19.0	3	28	0.7368	0.7040	0.0328
20.0	1	29	0.7632	0.7224	0.0407
21.0	2	31	0.8158	0.7396	0.0761
22.0	2	33	0.8684	0.7558	0.1126
25.0	1	34	0.8947	0.7985	0.0962
36.0	1	35	0.9211	0.9004	0.0206
39.0	1	36	0.9474	0.9178	0.0295
43.0	1	37	0.9737	0.9364	0.0373
54.0	1	38	1.0000	0.9686	0.0314
				Max=	0.1126

Fuente: Elaborado por el IMT

En la tabla 5.4, la máxima diferencia absoluta entre funciones de distribución acumuladas observada vs esperada es 0.1126. Este valor se compara con el valor crítico para la prueba K-S en una tabla para este propósito, que se puede encontrar en textos de estadística o de tablas estadísticas, resultando que el valor crítico es 0.226 para un nivel de confianza del 95%.

El resultado anterior permite entonces afirmar que los datos de llegadas de Aeroméxico analizado se comportan como una distribución Expo (0.641).

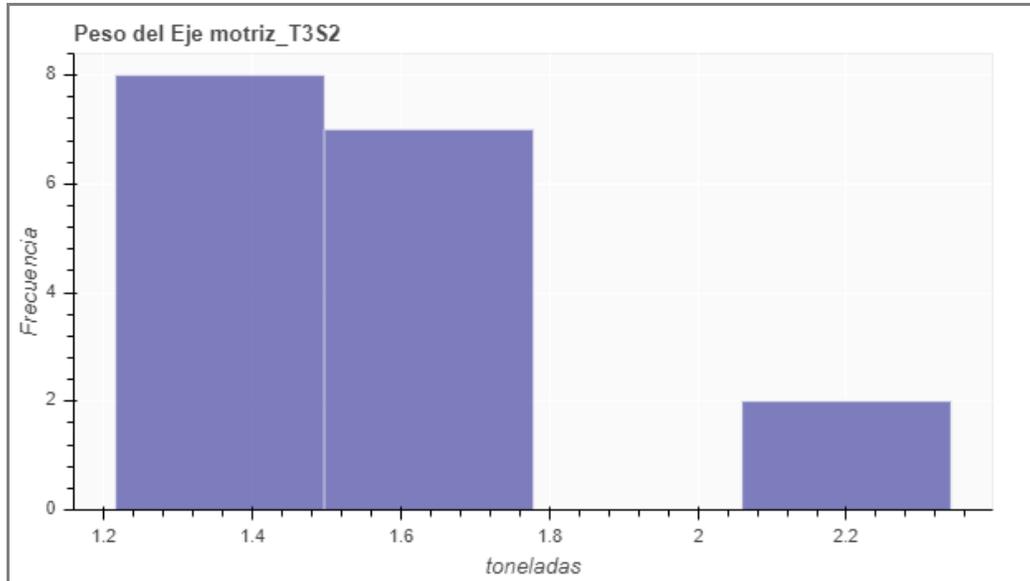
Ejemplo 5.3

La tabla 5.5 es una muestra de pesos del eje motriz (tons.) de camiones T3-S2 en transporte para el comercio exterior, registrados en la Estación 246 “Playas de Tijuana” de los estudios O-D realizados por la SCT en 2010. La figura 5.3 muestra el histograma correspondiente.

Tabla 5.5 Pesos del eje motriz. Camiones T3S2 Estación 246/2010

10.387	4.237	5.271	4.378
4.305	5.596	3.377	3.78
3.734	5.053	5.213	
5.606	4.646	5.856	
4.358	4.033	8.59	

Fuente: Elaborado por el IMT



Fuente: Elaborado por el IMT

**Figura 5.3 Peso del eje motriz. Camiones T3S2, Estación 246/2010
(elaboración propia con ABAL)**

El sesgo hacia la derecha del histograma sugiere una distribución Lognormal. Los parámetros para esta distribución se obtienen de la media y la varianza de los correspondientes valores del logaritmo natural de la muestra de datos, resultando: $\mu = 1.6050$, $\sigma = 0.2805$.

Los valores de toneladas de la muestra, así como los valores en orden creciente y decreciente, así como, las probabilidades acumuladas y los cálculos para la prueba Anderson-Darling se muestran en la tabla 5.6

Tabla 5.6 Detalle de la prueba A-D. Datos de peso del eje motriz T3S2

					A	B	C
n	Tons_eje	Y_j	Y_{17-j+1}	$2j - 1$	PAE_j	$1 - PAE_{17-j+1}$	$(2j - 1)[LnA + LnB]$
1	10.39	3.38	10.39	1	0.083	0.004	-7.918
2	4.31	3.73	8.59	3	0.153	0.026	-16.599
3	3.73	3.78	5.86	5	0.163	0.281	-15.405
4	5.61	4.03	5.61	7	0.227	0.336	-18.030
5	4.36	4.24	5.60	9	0.283	0.338	-21.122
6	4.24	4.31	5.27	11	0.302	0.419	-22.722
7	5.60	4.36	5.21	13	0.318	0.435	-25.737
8	5.05	4.38	5.05	15	0.324	0.479	-27.975
9	4.65	4.65	4.65	17	0.403	0.597	-24.221
10	4.03	5.05	4.38	19	0.521	0.676	-19.805
11	5.27	5.21	4.36	21	0.565	0.682	-20.006
12	3.38	5.27	4.31	23	0.581	0.698	-20.777
13	5.21	5.60	4.24	25	0.662	0.717	-18.633
14	5.86	5.61	4.03	27	0.664	0.773	-17.988
15	8.59	5.86	3.78	29	0.719	0.837	-14.745
16	4.38	8.59	3.73	31	0.974	0.847	-5.950
17	3.78	10.39	3.38	33	0.996	0.917	-3.015

Fuente: Elaborado por el IMT

El correspondiente estimador de la prueba resulta:

$$A_{17}^2 = - \left[17 + \frac{1}{17} \sum \text{Columna "C"} \right] = 0.6805$$

Este valor debe multiplicarse por el factor:

$$\left(1 + \frac{3}{4 \times 17} + \frac{9}{4 \times 17^2} \right)$$

ya que los parámetros μ y σ se estimaron a partir los datos; el estimador final es:

$$0.6805 \times \left(1 + \frac{3}{4 \times 17} + \frac{9}{4 \times 17^2} \right) \approx 0.72069$$

El valor crítico de la prueba Anderson-Darling con un nivel de significación $\alpha = 5\%$ para $n = 17$ se lee en la tabla 4.4 como 0.752, que resulta mayor al estimador final de la prueba y por tanto se puede concluir que los datos siguen una distribución Lognormal al nivel de significación del 5%.

6 Resultados y conclusiones

La necesidad de ajustar datos reales de carácter aleatorio a una distribución de probabilidad que los represente adecuadamente, teniendo un criterio estadístico de validación aparece con frecuencia en los trabajos de investigación, y de solución de problemas en el campo del transporte y en general en las ingenierías.

En la práctica no siempre se dispone de algún paquete comercial de estadística que permita realizar las pruebas de bondad de ajuste, y en el caso de tener acceso a alguno de ellos, no siempre están disponibles en los menús de esos paquetes todas las distribuciones que se desean probar.

La revisión de las distribuciones básicas utilizadas en los problemas de transporte e ingeniería mostradas en este trabajo, así como el resumen de los procedimientos de estimación de parámetros y la formulación de las pruebas de bondad de ajuste más comunes en la práctica: Ji-Cuadrada, Kolmogorov-Smirnov y Anderson-Darling, dan al lector los elementos para realizar todos los cálculos requeridos por cuenta propia.

Esto, por supuesto, requiere dedicar tiempo a codificar en algún lenguaje de programación o en el ambiente de Excel, y realizar pruebas piloto para asegurar que las respuestas obtenidas sean correctas.

La aplicación informática desarrollada en este proyecto se ha ocupado de todos los detalles prácticos del cálculo requerido en una prueba de bondad de ajuste.

Esto ha incluido la validación de los datos de entrada, los tamaños de muestra, la adecuación del tipo de dato a la prueba de bondad de ajuste que se desee utilizar, así como, los cálculos necesarios para estimar parámetros de las distribuciones que se quieren probar, y la evaluación misma de las técnicas de bondad de ajuste pertinentes: Ji-Cuadrada, Kolmogorov-Smirnov o Anderson-Darling según sea el tipo de dato cargado por el usuario.

En el desarrollo de la aplicación se utilizaron rutinas apropiadas disponibles en las librerías de Python, así como, un código cuidadosamente escrito para completar los cálculos que no se encuentran disponibles en estas librerías. Se desarrollaron también capacidades de graficación para obtener histogramas de los datos de entrada, funciones de masa de probabilidad para distribuciones discretas y curvas de densidad de probabilidad para distribuciones continuas, que permiten tener un reporte de resultados más completo para el usuario que utilice esta aplicación.

La aplicación se encuentra disponible en línea en: 10.34.1.95, y tiene la opción de exportar el reporte de la prueba de bondad de ajuste a fin de que los usuarios utilicen estos resultados en sus trabajos de investigación o de consultoría.

En esta primera versión beta de ABAL, se tiene una liga de contacto para que los usuarios opinen acerca del desempeño de la aplicación, o hagan consultas sobre las distribuciones que son de su interés. Con esta información que se colecte a medida que los usuarios utilicen ABAL, se podrán mejorar las ediciones futuras de la aplicación, y lograr un producto de evidente utilidad para la investigación y desarrollo tecnológico en el transporte y las ingenierías.

Bibliografía

STERN, Nicholas Herbert. The economics of climate change: the Stern review. Cambridge University press, 2007.

RIGGS, Glenn E., et al. Transport logistics systems and methods. U.S. Patent No 6,915,268, 5 Jul. 2005.

RIETVELD, Piet. Spatial economic impacts of transport infrastructure supply. Transportation Research Part A: Policy and Practice, 1994, vol. 28, no 4, p. 329-341.

KIM, Tschangho John; HAM, Heejoo; BOYCE, David E. Economic impacts of transportation network changes: Implementation of a combined transportation network and input-output model. Papers in Regional Science, 2002, vol. 81, no 2, p. 223-246.

Moreno, E. y Nieves, V. (2019). *Modelos de probabilidad en transporte e ingeniería. Usos comunes y ajuste de datos* Publicación Técnica No. 545. Instituto Mexicano del Transporte. Querétaro, México.

TIOBE - the software quality company (2019). *TIOBE Index for August 2019. August Headline: Silly season in the programming language world*. Recuperado de <https://www.tiobe.com/tiobe-index/>

COVANTEC (2018). *Características*. Recuperado de <https://entrenamiento-python-basico.readthedocs.io/es/latest/leccion1/caracteristicas.html>

WTForms (2010). *CSRF Protection*. Recuperado de <https://wtforms.readthedocs.io/en/stable/csrf.html>

Statistics in Python. (s.f). Recuperado de <http://www.scipy-lectures.org/packages/statistics/index.html>

ANACONDA (2019). *Anaconda Distribution*. Recuperado de <https://www.anaconda.com/distribution/>

ANACONDA.DOCUMENTATION (2019). *Anaconda Documentation*. Recuperado de <https://docs.anaconda.com/>

López, R.E. (2015, 27 de junio). Probabilidad y estadística con Python. Mensaje dirigido a <https://relopezbriega.github.io/blog/2015/06/27/probabilidad-y-estadistica-con-python/>



Km 12+000 Carretera Estatal 431 "El Colorado-Galindo"
Parque Tecnológico San Fandila
Mpio. Pedro Escobedo, Querétaro, México
CP 76703
Tel +52 (442) 216 9777 ext. 2610
Fax +52 (442) 216 9671

publicaciones@imt.mx

<http://www.imt.mx/>