ESCUELA UNIVERSITARIA POLITÉCNICA
DE LA ALMUNIA DE DOÑA GODINA (ZARAGOZA)

MEMORIA

[APLICACCIÓN WEB DE INVENTARIOS CON DEMANDA ESTOCASTICA]

[425.16.106]

Autor: [MARIANO CIUTAD BUETAS]

Director: [Luis Mariano Esteban Escaño]

Fecha: 28/06/2016

INDICE DE CONTENIDO

INDICE DE ILUSTRACIONES

1. RESUMEN

Este trabajo está enfocado al cálculo del tamaño de un lote que debemos solicitar dentro del ámbito de la optimización de inventarios, teniendo en cuenta que dependemos de una demanda aleatoria y unos costes.

Esto nos conduce a tener que deducir que tipo de distribución de la demanda sigue una muestra, y los costes relacionados con la misma. Una vez establecidas las bases pasamos al siguiente punto que se centra en crear un programa 'OPTINV', que realice a modo de calculadora exigiendo unos inputs (muestra de la demanda 'Excel', costes por faltantes, costes de almacén, costes de producción), también decidimos establecer el cálculo de un nivel de stock, y el nivel de inventario inicial, por lo que también establecemos dichas entradas en el programa. Posteriormente el programa realiza los cálculos preestablecidos y nos dirige a un ajuste de la demanda, la cual el programa nos muestra una tabla con los errores acumulados, y en esta misma debemos seleccionar el tipo de distribución que más se nos ajuste.

Finalmente nos muestra los datos del lote a pedir, stock de seguridad, lote total a pedir, acompañado de un gráfico.

2. ABSTRACT

This work focuses on the calculation of the size of a batch that must apply in the field of inventory optimization, considering that depend on a random demand and costs.

This leads us to have to deduce what kind of distribution of demand is a sample, and the costs related thereto. Once we established the foundation spent the next point that focuses on creating a 'OPTINV' program, which perform as a calculator demanding some inputs (sample demand 'Excel' costs for missing, warehouse costs, production costs), also we decided to establish the calculation of a stock level, and the level of initial inventory, so we also set these entries in the program. Subsequently, the program performs the preset calculations and directs us to an adjustment in demand, which the program shows a table with accumulated errors, and the same must select the type of distribution that we are more fit.

Finally it shows batch data to ask, safety stock, total lot to ask, accompanied by a graphic.

3. INTRODUCCIÓN

Este trabajo de fin de grado surge al conocer el sector de la logística y producción, este sector es una parte fundamental dentro de las empresas ya que es el organismo que regula la circulación de productos. Al conocer en la carrera asignaturas como investigación operativa y la parte de esta en la que se especializa en Teoría de Inventarios, observamos que este sector nos proporciona una base teórica para la minimización de costes, uno de los objetivos principales dentro de las empresas, y de aquí surge el objetivo principal del trabajo que es minimizar los costes que están relacionados con la ordenación de inventarios.

Para perfilar la demanda aleatoria dependeremos de un gran abanico de opciones de tipos de distribuciones dotados de ecuaciones complejas, para calcular el nivel optimo del lote a pedir, debemos deducir que tipo de distribución ya que se trata de un distribución aleatoria.

Como hemos podido observar en el anterior párrafo, hay un buen número de ecuaciones además de ser de compleja resolución, haciendo hincapié en la asignatura de Investigación operativa y observar la herramienta de cálculo estadístico "R" podemos entender que este programa nos facilita el cálculo de un gran número de las ecuaciones antes mostrada. Dentro de este ámbito encontramos un paquete (explicado posteriormente) que nos facilita la creación de un app la cual está relacionada con la herramienta de "R-studio" para la resolución del conjunto de problemas anteriormente mostrados.

Tras ver el funcionamiento de dicho programa nos surge una duda, porque como hemos comentado es una herramienta estadística y no todo el mundo sabrá usarla, por ello establecemos la prioridad del trabajo de fin de grado, que será crear un app funcional y sencilla de utilizar. Esta idea proviene, al pensar como consumidores de este programa a las empresas y estas no tener en su mayoría un departamento de estadística.

La motivación de esta parte del proyecto está directamente relacionada en crear una herramienta de trabajo fácil de usar y sin tener grandes conocimientos tanto de estadística como matemáticos.

Como hemos ido observando, el objetivo principal del TFG es desarrollar una herramienta práctica para el cálculo de niveles óptimos de inventario, así que el proyecto incluye dos partes, por un lado una introducción sobre teoría de inventarios y sus bases estadísticas y matemáticas, y por otro lado el estudio de una herramienta que nos facilite la creación de un programa para optimizar el inventario de una empresa. Para la creación de dicho programa elegimos Rstudio, ya que se trata de un herramienta estadística, que tiene una fácil programación y dispone de paquetes para resolver problemas de uso específico, también contamos con un foro de ayuda así como también con tutoriales y demás apoyos técnicos dentro de su web, y por último, es una aplicación de software libre lo que favorece su selección.

Para el funcionamiento del programa necesitaremos una serie de valores de entrada o inputs que determinaran el menor coste, los cuales están definidos en el tema de Teoría de Inventarios.

- Costes de producción.
- Costes de mantenimiento.
- Costes de almacenaje.
- Costes por faltantes.
- % de stock de seguridad.
- Nivel de inventario inicial.

Dentro del tema sabiendo esto y qué tipo de distribución de demanda tiene el producto, a partir de una muestra podemos calcular el lote optimo a pedir, pero como hemos comentado anteriormente creemos que la mayoría de las PYMES, no cuentan con un departamento de estadística o matemáticas que les ayuden a calcular dicha función, por ello decidimos que para facilitar el trabajo y tener un abanico mayor a la hora de oportunidades debíamos introducir en el programa un apartado que pudiese calcular que tipo de distribución de demanda tiene una muestra, lo que fomento un estudio sobre modelos estadísticos para encontrar la mejor manera de calcular dicha función.

Para ajustar la distribución de la demanda necesitamos los datos de demanda durante un período, dichos datos serán una entrada en el programa que aparecerá como una tabla Excel ya que la mayoría de la empresas trabajan con ellas, en la cual estuviesen reflejados las demandas en un periodo de servicio de sus correspondiente proveedores, con esto queremos decir que decidimos no establecer también el cuándo reponer ya que la mayoría de la empresas tienen proveedores que sirven dentro de unas fecha y unos horarios y sería difícil cambiarlos, así definimos otra entrada de datos en el programa.

- Archivo Excel.

La dificultad del proyecto reside en establecer unas funciones que determinen a qué tipo de función de distribución se asemeja una muestra.

Una vez tenemos definidas las entradas del programa establecemos que es lo que queremos como salida del mismo.

- Tamaño del tote total.
- Lote a pedir.
- Stock de seguridad.

Finalmente decidimos establecer un nombre a dicho programa, en nuestro caso optamos por el siguiente OPTINV.

En definitiva se trata de una herramienta que está enfocada a resolver los posibles problemas que pueda tener una empresa a la hora de ir calculando los niveles de inventarios.

4. DESARROLLO

4.1. OBJETIVOS GLOBALES:

Antes de comenzar a tratar con el conjunto del proyecto decidimos establecer unos objetivos cuantificables, que nos dieran la posibilidad de ir atando cabos para conseguir el objetivo principal, crear un programa capaz de calcular inventarios.

1- Formular un modelo matemático que describa el comportamiento del sistema de inventarios.
2- Derivar una política óptima de inventarios con respecto al modelo.
3- Mantener un registro de los niveles de inventario y señalar cuando y cuanto conviene restablecer.

Como podemos observar estos objetivos encajan dentro del ámbito productivo como el enfocado a la logística, para ello definimos como afectarían a cada sector, teniendo en cuenta que los costes son constantes la variable a tener en cuenta sería el tipo de demanda seleccionada.

1- Sector productivo: el tipo de producción de cada empresa puede variar entre, ser una producción constante, para la cual necesitaríamos conocer el número de unidades requeridas, o la producción puede estar fragmentada creando una producción racheada.

   Para este sector como hemos observado encontramos unos objetivos preestablecidos y que señalaran en la mayoría de las veces a una producción con demanda determinista.
2- Sector logístico: a diferencia del sector productivo la demanda que este requiere es casi siempre aleatoria, lo que nos da un planteamiento de los modelos matemáticos más complejo.

Para dicha reordenación de inventarios es muy importante conocer la demanda, este parámetro como iremos viendo es el pilar del trabajo.

Leído lo anterior nos damos cuenta de que sabemos lo que queremos pero ahora debemos establecer como lo queremos.

Para ello establecemos necesidades:

1. Precisamos un archivo que nos recoja toda la información de la demanda, y suponemos que la gran mayoría de las empresas usan Excel como programa de archivo de datos.
   Establecemos así lo que sería la entrada de datos.
   (a) **Entrada de datos será un archivo Excel.**

2. También debemos establecer otra entrada de datos para los costes, por lo que tendremos las siguientes ventanas para introducir los datos:
   (a) **Costes de mantenimiento**
   (b) **Costes por faltantes**
   (c) **Costes de producción**

3. Decidimos establecer una entrada de datos como el nivel de stock que tenemos en inventario.
   (a) **Stock a principio de periodo**

Como podemos observar tenemos todas las entradas ya predeterminadas y ahora tenemos que decidir qué tipo de salida queremos.

4. El programa queremos que nos muestre en pantalla:
   (a) **Lote óptimo total**
   (b) **Lote a pedir**
   (c) **Stock de seguridad**

El párrafo anterior tiene aplicado el estilo TFG_Título 2

4.2. ESTUDIO TEÓRICO DE LAS BASES DEL PROBLEMA:

Una vez establecemos los criterios a seguir dentro del problema que se nos plantea, debemos realizar un estudio teórico sobre las posibles soluciones y métodos de optimización.

4.2.1. TEORÍA DE INVENTARIOS:

Los inventarios prevalecen en el mundo de los negocios. Mantener inventarios es necesario para las compañías que tratan con productos físicos, como fabricantes, distribuidores y comerciantes. Por ejemplo, los fabricantes necesitan inventarios de materiales requeridos para la manufactura de productos. También deben almacenar productos terminados en espera de ser enviados. De manera similar, tanto los distribuidores como las tiendas deben mantener inventarios de bienes disponibles cuando los consumidores los necesiten. Reducir los costos de almacenamiento evitando inventarios innecesariamente grandes puede mejorar la competitividad de cualquier empresa. Algunas compañías japonesas han sido pioneras en la introducción de los "sistemas de inventarios justo a tiempo", un sistema que hace hincapié en la planeación y programación para que los materiales necesarios lleguen "justo a tiempo" para su uso. La aplicación de técnicas de la investigación de operaciones en esta área (administración científica de los inventarios) proporciona una herramienta poderosa para lograr una ventaja competitiva.

4.2.1.1. Definición de inventario:

Inventario son las existencias de cualquier artículo o recurso utilizado en una organización. Un sistema de inventario es la serie de políticas y controles que monitorean los niveles de inventario y determinan los niveles que se deben mantener, el momento en que las existencias se deben reponer y el tamaño que deben tener los pedidos. El inventario en el sector manufacturero se clasifica típicamente en materias primas, productos terminados, partes componentes, suministros y trabajo en proceso. En el sector servicios, el inventario se refiere generalmente a los bienes tangibles que van a venderse y a los suministros necesarios para administrar el servicio. El objetivo básico del análisis de inventario en el sector manufacturero y en los servicios de mantenimiento de las existencias es especificar (1) cuando se deben ordenar los artículos y (2) que tan grande debe ser el pedido.

A la hora de mantener un inventario, además de saber qué tipo de demanda sigue, tenemos que especificar los costes que influyen sobre este, estos serán constantes ya que dependerán de otros departamentos de la empresa como producción, mantenimiento, almacén, distribución, etc.

4.2.1.2. Costos de inventario:

1. Costo de ordenar: Estos se refieren a los costos administrativos y de oficina para elaborar la orden de compra o de producción. Los costos de las órdenes incluyen todos los detalles, tales como contar los artículos y calcular las cantidades de órdenes. Los costos asociados con el mantenimiento del sistema necesario para rastrear las órdenes están también incluidos en estos costos.
2. Costo de almacenamiento: Esta categoría incluye los costos de las instalaciones de almacenamiento, el manejo, el seguro, hurto, la rotura, la obsolescencia, la depreciación, los impuestos, y el costo de oportunidad del material. Obviamente, los altos costos de mantenimiento tienden a favorecer unos bajos niveles de inventarios y la reposición frecuente.
3. Costo de producción: La fabricación de cada producto diferente implica obtener los materiales necesarios, arreglar la preparación del equipo especifico, diligenciar los documentos requeridos, cargar de manera apropiada el tiempo, y los materiales, y desalojar los anteriores suministros de material. Si no hubieran costos o pérdida de tiempo en cambiar de un producto a otro, se producirían muchos lotes pequeños. Esto reduciría los niveles de inventario con el resultante ahorro en el costo. Un desafío en la actualidad es tratar de reducir estos costos de preparación para permitir unos tamaños de lotes más pequeños.
4. Costos por faltantes: cuando las existencias de un artículo están agotadas, los pedidos de ese artículo deben esperar hasta que estas se repongan o cancelarse. Existe una transacción entre llevar las existencias para satisfacer la demanda y los costos resultantes del agotamiento de las mismas. Este equilibrio es difícil de lograr, porque no es posible calcular las utilidades perdidas, los efectos de perder clientes o las sanciones por retraso.

Establecer la cantidad correcta que debe pedirse a los proveedores, o el tamaño de los lotes presentados a las instalaciones productivas de una firma, implica una búsqueda del costo total mínimo resultante de los efectos combinados de cuatro costos individuales: los costos de mantenimiento, los costos de preparación, los costos de los pedidos, y los costos de los faltantes.

4.2.1.3. Diferencias entre demanda dependiente y demanda aleatoria:

- ANÁLISIS CON DEMANDA DEPENDIENTE:

En el mantenimiento de inventarios, el conocimiento de la demanda resulta fundamental. Si la demanda es determinista, cuándo y cuánto ordenar es el resultado de tener en cuenta la posibilidad de tener que asumir costos fijos en la orden, admitir la posibilidad de faltantes o tener descuentos en el coste de producir u ordenar.

Para cada caso, obtenemos una fórmula que nos determina el lote óptimo Q o si es necesario la cantidad de faltantes que podemos asumir.

Estos modelos deterministas están enfocados a casos de demanda particulares, en los cuales la demanda se conoce y se trabaja con unos objetivos predeterminados, haciendo que los costes sean rápidamente calculados, este tipo de demanda lleva unos periodos y lotes prácticamente lineales y la mayoría de las veces están destinados a la producción bajo contrato.

- ANÁLISIS CON DEMANDA ESTOCÁSTICA:

En estadística, y específicamente en la teoría de la probabilidad, un proceso estocástico es un concepto matemático que sirve para tratar con magnitudes aleatorias que varían con el tiempo, o más exactamente para caracterizar una sucesión de variables aleatorias(estocásticas) que evolucionan en función de otra variable, generalmente el tiempo. Cada una de las variables aleatorias del proceso tiene su propia función de distribución de probabilidad y pueden o no, estar correlacionadas entre ellas.

Cada variable o conjunto de variables sometidas a influencias o efectos aleatorios constituye un proceso estocástico. Un proceso estocástico X_t puede entenderse como una familia uniparamétrica de variables aleatorias indexadas mediante el tiempo t. Los procesos estocásticos permiten tratar procesos dinámicos en los que hay cierta aleatoriedad.

Si la demanda es estocástica, esto quiere decir que la demanda de un producto tendrá una distribución aleatoria. Debido a que la mayoría de empresas trabajan con demandas aleatorias decidimos crear un Trabajo de fin de grado que se centrara en solucionar este tipo de problemas, por eso recurrimos a la estadística como modelo matemático.

Como vemos los problemas que se puedan plantear dentro de este trabajo están directamente relacionados con la estadística. Para la resolución de este tipo de problemas no tenemos una única fórmula, ya que tenemos que exigir una serie de datos al cliente, los cuales determinaran que tipo de cálculo algoritmo el cual nos determinara el valor a ordenar, y periodos de orden.

El párrafo anterior tiene aplicado el estilo TFG_Título 3

Para escribir listas de datos deberás de emplear el estilo TFG_Viñetas. A continuación tienes un ejemplo:

- Elemento 1
- Elemento 2
- Elemento 3

Para insertar notas al pie, empleamos el comando "insertar notas al pie" con las especificaciones que por defecto incluye dicho comando. Se encuentra en la ficha referencias, grupo notas al pie.

4.2.2. DEMANDA ESTOCASTICA

El párrafo anterior tiene aplicado el estilo TFG_Título 2

Se denomina estocástico (del latín stochasticus, que a su vez procede del griego στοχαστικός, "hábil en conjeturar")1 al sistema cuyo comportamiento es intrínsecamente no determinista. Un proceso estocástico es aquel cuyo comportamiento es no determinista, en la medida que el subsiguiente estado del sistema está determinado tanto por las acciones predecibles del proceso como por elementos aleatorios. No obstante, de acuerdo con M. Kac2 y E. Nelson,3 cualquier desarrollo temporal (sea determinista o esencialmente probabilístico) que pueda ser analizable en términos de probabilidad merece ser denominado como un proceso estocástico.

4.2.3. INTRODUCCIÓN ESTADÍSTICA SOBRE DISTRIBUCIONES ALEATORIAS

Para escribir listas de datos deberás de emplear el estilo TFG_Viñetas. A continuación tienes un ejemplo:

- Elemento 1
- Elemento 2
- Elemento 3

Para insertar notas al pie, empleamos el comando "insertar notas al pie" con las especificaciones que por defecto incluye dicho comando. Se encuentra en la ficha referencias, grupo notas al pie.

Otro contenido…

Para describir la demanda aleatoria tendremos que saber qué tipo de demanda siguen los datos teniendo en cuenta que hay dos parámetros, función de distribución de probabilidad y función de densidad de probabilidad.

Estos son los parámetros a tener en cuenta para el calculo.

Distribución de demanda aleatoria: la distribución de probabilidad de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable aleatoria la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos los sucesos, cada uno de los sucesos es el rango de valores de la variable aleatoria.

Todas las funciones que definen una región en la cual la probabilidad de que ocurra un suceso está definida por una función de densidad y una función de distribución de la probabilidad.

Con la siguiente nomenclatura matemática:

Dada una variable aleatoria X, su función de distribución, $F_X(x)$, es:

$$F_X(x) = \mathrm{Prob}(X \leq x) = \mu_P\{\omega \in \Omega | X(\omega) \leq x\}$$

Por simplicidad, cuando no hay lugar a confusión, suele omitirse el subíndice X y se escribe, simplemente, $F(x)$. Donde en la fórmula anterior:

Prob, Es la probabilidad definida sobre un espacio de probabilidad y una medida unitaria sobre el espacio muestral.

μ_P Es la medida sobre la σ-álgebra de conjuntos asociada al espacio de probabilidad.

Ω Es el espacio muestral, o conjunto de todos los posibles sucesos aleatorios, sobre el que se define el espacio de probabilidad en cuestión.

$X : \Omega \to \mathbb{R}$ Es la variable aleatoria en cuestión, es decir, una función definida sobre el espacio muestral a los números reales.

Función de densidad de probabilidad: En la teoría de la probabilidad, la función de densidad de probabilidad, función de densidad, o, simplemente, densidad de una variable aleatoria continua describe la probabilidad relativa según la cual dicha aleatoriedad tomará determinado valor.

La probabilidad de que la variable aleatoria caiga en una región específica del espacio de posibilidades estará dada por la integral de la densidad de esta variable entre uno y otro límite de dicha región.

La función de densidad de probabilidad: describe el área donde puede ocurrir un suceso, como mostramos en la siguiente imagen:

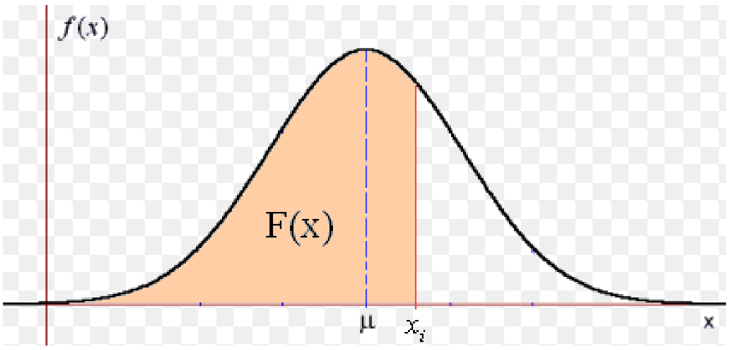


Ilustración 1FUNCIÓN DE DESNSIDAD DE PROBABILIDAD

La función de densidad de probabilidad (FDP o PDF en inglés) es no-negativa a lo largo de todo su dominio y su integral sobre todo el espacio es de valor unitario.

Con la siguiente nomenclatura matemática:

Una función de densidad de probabilidad caracteriza el comportamiento probable de una población en tanto especifica la posibilidad relativa de que una variable aleatoria continua X tome un valor cercano a x.

Una variable aleatoria X tiene densidad f, siendo f una función no-negativa integrable de Lebesgue, si:

$$\mathrm{P}[a \leq X \leq b] = \int_a^b f(x)\,\mathrm{d}x.$$

Por lo tanto, si F es la función de distribución acumulativa de X, entonces:

$$F(x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u,$$

Y (si f es continua en x)

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x} F(x).$$

Intuitivamente, puede considerarse f(x) dx como la probabilidad de X de caer en el intervalo infinitesimal [x, x + dx].

4.2.3.1. TIPOS DE DISTRIBUCIÓN

Los tipos de distribución de probabilidad, siguen unos parámetros que determinan unas aéreas en las cuales los sucesos son más probables que ocurran y otras con menor probabilidad, como podemos entender existen infinidad de modelos de distribuciones. En nuestro caso hemos elegido los modelos que podrían aproximarse como solución del problema.

Estas distribuciones se emplearon en el estudio de fenómenos aleatorios en disciplinas como la ingeniería y las ciencias aplicadas o bien los negocios y la economía.

- DISTRIBUCIÓN UNIFORME:

En teoría de probabilidad y estadística, la distribución uniforme continua es una familia de distribuciones de probabilidad para variables aleatorias continuas, tales que cada miembro de la familia, todos los intervalos de igual longitud en la distribución en su rango son igualmente probables. El dominio está definido por dos parámetros, a y b, que son sus valores mínimo y máximo. La distribución es a menudo escrita en forma abreviada como U(a,b).

Función de densidad de probabilidad

La función de densidad de probabilidad de la distribución uniforme continua es:

$$f(x)=\begin{cases}\frac{1}{b-a} & \text{para } a \le x \le b, \\ 0 & \text{para } x < a \text{ o } x > b,\end{cases}$$

Aquí mostramos como queda la grafica de dicho comportamiento:

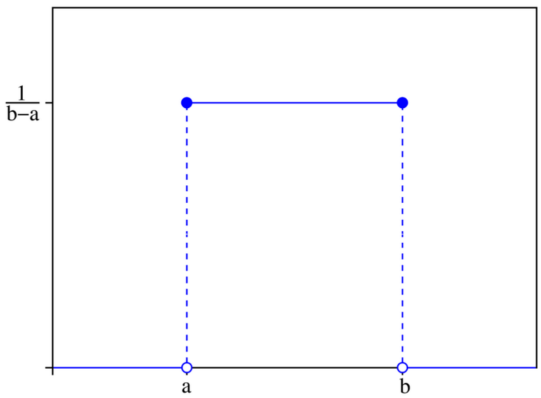


Ilustración 2 FUNCIÓN DE DENSIDAD DE PROBABILIDAD UNIFORME

Los valores en los dos extremos a y b no son por lo general importantes porque no afectan el valor de las integrales de f(x) dx sobre el intervalo, ni de x f(x) dx o expresiones similares. A veces se elige que sean cero, y a veces se los elige con el valor 1/(b − a). Este último resulta apropiado en el contexto de estimación por el método de máxima verosimilitud. En el contexto del análisis de Fourier, se puede elegir que el valor de f(a) ó f(b) sean 1/(2(b − a)), para que entonces la transformada inversa de muchas transformadas integrales de esta función uniforme resulten en la función inicial, de otra forma la función que se obtiene sería igual "en casi todo punto", o sea excepto en un conjunto de puntos con medida nula. También, de esta forma resulta consistente con la función signo que no posee dicha ambigüedad.

La función de distribución de probabilidad es:

$$F(x) = \begin{cases} 0 & \text{para } x < a \\ \frac{x-a}{b-a} & \text{para } a \leq x < b \\ 1 & \text{para } x \geq b \end{cases}$$

Con la consiguiente gráfica:

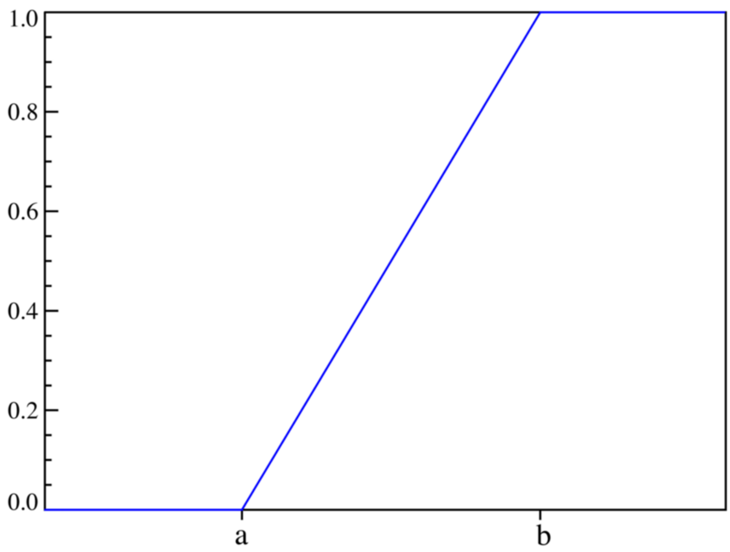


Ilustración 3 FUNCIÓ DE DISTRIBUCIÓN DE PROBABILIDAD UNIFORME

La distribución normal o Gausiana es indudablemente la más importante y una de las más usadas como distribución continúa.

- <u>DISTRIBUCION NORMAL:</u>

Los parámetros de la distribución normal son (μ, σ) y además determinan de manera completa la función de densidad de probabilidad. Como se verá posteriormente, estos parámetros son la media y la desviación estándar de X, respectivamente.

La distribución normal también aparece en muchas áreas de la propia estadística. Por ejemplo, la distribución muestral de las medias muestrales es aproximadamente normal, cuando la distribución de la población de la cual se extrae la muestra no es normal.1 Además, la distribución normal maximiza la entropía entre todas las distribuciones con media y varianza conocidas, lo cual la convierte en la elección natural de la distribución subyacente a una lista de datos resumidos en términos de media muestral y varianza. La distribución normal es la más extendida en estadística y muchos test estadísticos están basados en una "normalidad" más o menos justificada de la variable aleatoria bajo estudio.

En probabilidad, la distribución normal aparece como el límite de varias distribuciones de probabilidad continuas y discretas.

La función de distribución de la distribución normal está definida como sigue:

$$\begin{aligned}\Phi_{\mu,\sigma^2}(x) &= \int_{-\infty}^{x} \varphi_{\mu,\sigma^2}(u)\,du \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}}\,du, \quad x \in \mathbb{R}\end{aligned}$$

Por tanto, la función de distribución de la normal estándar es:

$$\Phi(x) = \Phi_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du, \quad x \in \mathbb{R}.$$

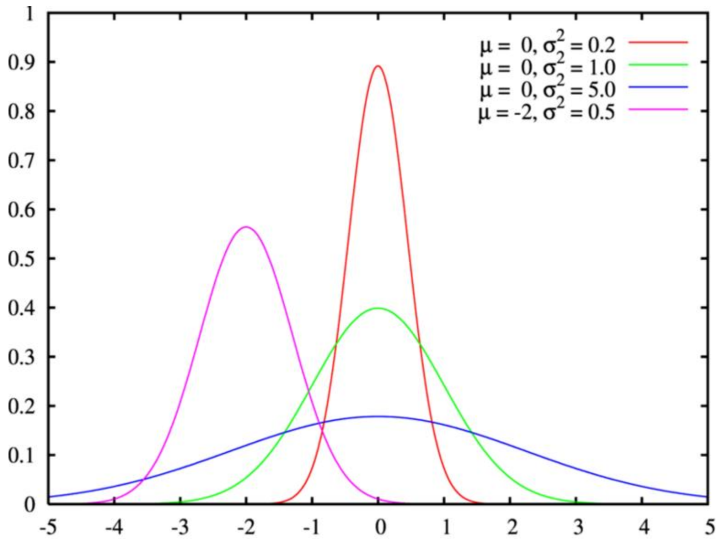


Ilustración 4 FUNCIÓN DE DENSIDAD DE PROBABILIDAD NORMAL

Función de distribución de probabilidad:

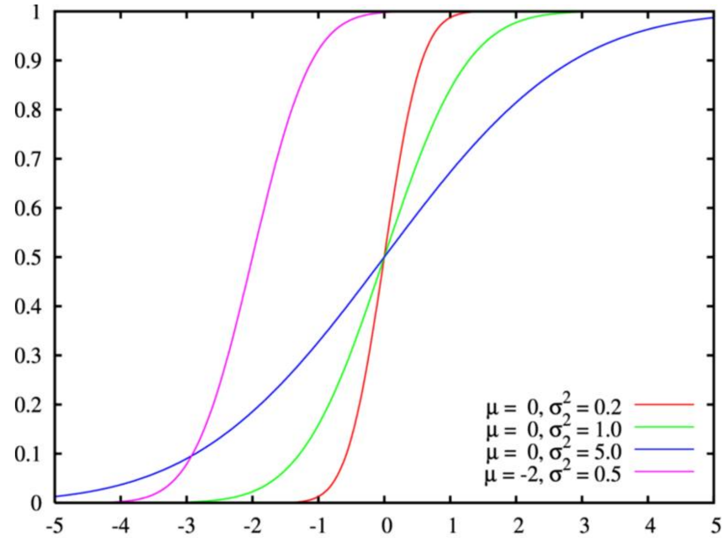


Ilustración 5 FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD NORMAL

Esta función de distribución puede expresarse en términos de una función especial llamada función error de la siguiente forma:

$$\Phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right], \quad x \in \mathbb{R},$$

y la propia función de distribución puede, por consiguiente, expresarse así:

$$\Phi_{\mu,\sigma^2}(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right], \quad x \in \mathbb{R}.$$

El complemento de la función de distribución de la normal estándar, $1 - \Phi(x)$, se denota con frecuencia $Q(x)$, y es referida, a veces, como simplemente función Q, especialmente en textos de ingeniería.5 6 Esto representa la cola de probabilidad de la distribución gaussiana. También se usan ocasionalmente otras definiciones de la función Q, las cuales son todas ellas transformaciones simples de Φ.7

La inversa de la función de distribución de la normal estándar (función cuantil) puede expresarse en términos de la inversa de la función de error:

$$\Phi^{-1}(p) = \sqrt{2}\ \operatorname{erf}^{-1}(2p-1), \quad p \in (0,1),$$

y la inversa de la función de distribución puede, por consiguiente, expresarse como:

$$\Phi_{\mu,\sigma^2}^{-1}(p) = \mu + \sigma\Phi^{-1}(p) = \mu + \sigma\sqrt{2}\ \operatorname{erf}^{-1}(2p-1), \quad p \in (0,1).$$

Esta función cuantil se llama a veces la función probit. No hay una primitiva elemental para la función probit. Esto no quiere decir meramente que no se conoce, sino que se ha probado la inexistencia de tal función. Existen varios métodos exactos para aproximar la función cuantil mediante la distribución normal (véase función cuantil).

Los valores Φ(x) pueden aproximarse con mucha precisión por distintos métodos, tales como integración numérica, series de Taylor, series asintóticas y fracciones continuas.

Límite inferior y superior estricto para la función de distribución

Para grandes valores de x la función de distribución de la normal estándar $\Phi(x)$ es muy próxima a 1 y $\Phi(-x) = 1 - \Phi(x)$ está muy cerca de 0. Los límites elementales

$$\frac{x}{1+x^2}\varphi(x) < 1 - \Phi(x) < \frac{\varphi(x)}{x}, \qquad x > 0,$$

en términos de la densidad φ son útiles.

Usando el cambio de variable v = u²/2, el límite superior se obtiene como sigue:

$$\begin{aligned} 1 - \Phi(x) &= \int_x^\infty \varphi(u)\, du \\ &< \int_x^\infty \frac{u}{x}\varphi(u)\, du = \int_{x^2/2}^\infty \frac{e^{-v}}{x\sqrt{2\pi}}\, dv = -\frac{e^{-v}}{x\sqrt{2\pi}}\bigg|_{x^2/2}^\infty = \frac{\varphi(x)}{x}. \end{aligned}$$

De forma similar, usando $\varphi'(u) = -u\,\varphi(u)$ y la regla del cociente,

[425.16.106]

Desarrollo

$$\begin{aligned}\Big(1+\frac{1}{x^2}\Big)(1-\Phi(x)) &= \Big(1+\frac{1}{x^2}\Big)\int_x^\infty \varphi(u)\,du \\ &= \int_x^\infty \Big(1+\frac{1}{x^2}\Big)\varphi(u)\,du \\ &> \int_x^\infty \Big(1+\frac{1}{u^2}\Big)\varphi(u)\,du = -\frac{\varphi(u)}{u}\Big|_x^\infty = \frac{\varphi(x)}{x}.\end{aligned}$$

Resolviendo para $1-\Phi(x)$ proporciona el límite inferior.

Siendo "erf" la función error como explicamos a continuación:

En matemáticas, la función error (también conocida como función error de Gauss) es una función especial (no elemental) que se utiliza en el campo de la probabilidad, la estadística y las ecuaciones diferenciales parciales. La función queda definida por la expresión:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,\mathrm{d}t$$

La función error complementaria, llamada erfc, se define a partir de la función error:

$$\operatorname{erfc}(x) = 1-\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}\,\mathrm{d}t$$

- <u>DISTRIBUCIÓN LOG NORMAL:</u>

En probabilidades y estadísticas, la distribución normal logarítmica es una distribución de probabilidad de una variable aleatoria cuyo logaritmo está normalmente distribuido. Es decir, si X es una variable aleatoria con una distribución normal, entonces exp(X) tiene una distribución log-normal.

La base de una función logarítmica no es importante, ya que loga X está distribuida normalmente si y sólo si logb X está distribuida normalmente, sólo se diferencian en un factor constante.

Log-normal también se escribe log normal o lognormal.

Una variable puede ser modelada como log-normal si puede ser considerada como un producto multiplicativo de muchos pequeños factores independientes. Un ejemplo típico es un retorno a largo plazo de una inversión: puede considerarse como un producto de muchos retornos diarios.

La distribución log-normal tiende a la función densidad de probabilidad

Función de densidad de probabilidad:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln(x)-\mu)^2/2\sigma^2}$$

Dicha expresión muestra las siguientes características gráficas como vemos en la siguiente imagen:

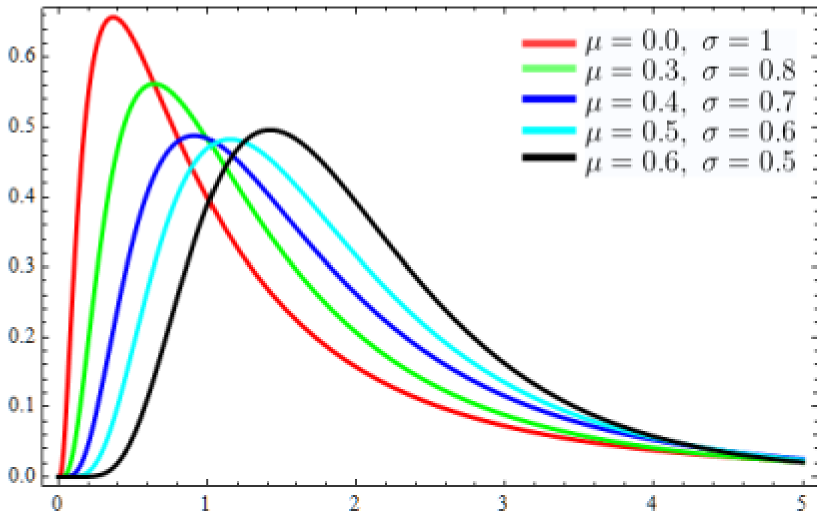


Ilustración 6 FUNCIÓN DE DENSIDAD DE PROBABILIDAD LOG-NORMAL

Relación con media y la desviación estándar geométrica

La distribución log-normal, la media geométrica, y la desviación estándar geométrica están relacionadas. En este caso, la media geométrica es igual a $\exp(\mu)$ y la desviación estándar geométrica es igual a $\exp(\sigma)$.

Si una muestra de datos determina que proviene de una población distribuida siguiendo una distribución log-normal, la media geométrica de la desviación estándar geométrica puede utilizarse para estimar los intervalos de confianza tal como la media aritmética y la desviación estándar se usan para estimar los intervalos de confianza para un dato distribuido normalmente.

Límite de intervalo de confianza	log	geométrica
3σ límite inferior	$\mu -$	$\mu_{\text{geo}} /$
2σ límite inferior	$\mu -$	$\mu_{\text{geo}} /$
1σ límite inferior	$\mu -$	$\mu_{\text{geo}} /$
1σ límite superior	$\mu +$	$\mu_{\text{geo}} \sigma$
2σ límite superior	$\mu +$	$\mu_{\text{geo}} \sigma$
3σ límite superior	$\mu +$	$\mu_{\text{geo}} \sigma$

Donde la media geométrica $\mu_{\text{geo}} = \exp(\mu)$ y la desviación estándar geométrica $\sigma_{\text{geo}} = \exp(\sigma)$

- <u>DISTRIBUCIÓN DE WEIBULL:</u>

La distribución de Weibull fue establecida por el físico Suizo del mismo nombre, quien demostró, con base en una evidencia empírica, que el esfuerzo de al que se somete los materiales puede modelarse de manera adecuada mediante el empleo de esta distribución. Esta distribución está caracterizada en los últimos años por determinar tiempo-falla, con el objetivo de ampliar la variedad de componentes mecánico electrónicos.

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

Función de densidad de probabilidad:

$$f(x;\alpha,\theta) = \begin{cases} \alpha * ((x^{\alpha-1}) * \exp\left[-\left(\frac{x}{\theta}\right)^{\alpha}\right])/(\theta^{\alpha}), & x > 0;\ \alpha,\theta > 0, \\ 0, & \textit{Para cualquier otro caso} \end{cases}$$

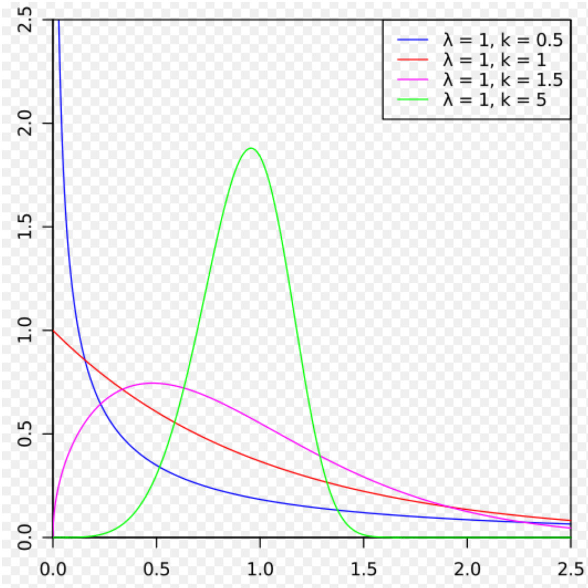


Ilustración 7 FUNCIÓN DE DENSIDAD DE PROBABILIDAD WEIBULL

Siendo λ=α y k= θ.

La distribución de Weibull es una familia de distribuciones que dependen de dos parámetros: de la forma α y de la escala θ. Se puede introducir un parámetro adicional al reemplazar la variable aleatoria de Weibull X por X-a, donde 'a'el parámetro que define un valor umbral o tiempo de garantía.

Función de distribución de probabilidad:

$$f(x;\alpha,\theta) = 1 - e^{-\left(\frac{x}{\alpha}\right)^{\theta}}$$

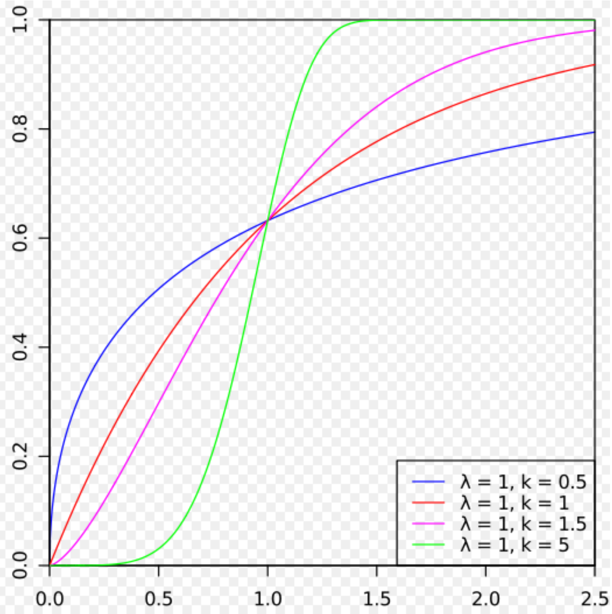


Ilustración 8 FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD WEIBULL

Siendo λ=α y k= θ.

- <u>DISTRIBUCIÓN LOGÍSTICA:</u>

En teoría de la probabilidad y estadística, la distribución logística es una distribución de probabilidad continua cuya función de distribución es la función logística, que aparece en el contexto de la regresión logística y determinados tipos de redes neuronales. Se parece a la distribución normal en su forma, pero tiene colas más pesadas (y, por lo tanto, menor curtosis).

La distribución logística recibe su nombre de su función de distribución, que pertenece a la familia de las funciones logísticas:

$$F(x;\mu,s) = \frac{1}{1+e^{-(x-\mu)/s}}$$

$$= \frac{1}{2} + \frac{1}{2}\,\tanh\!\left(\frac{x-\mu}{2\,s}\right).$$

Su función de distribución de la probabilidad queda mostrada en la siguiente imagen:

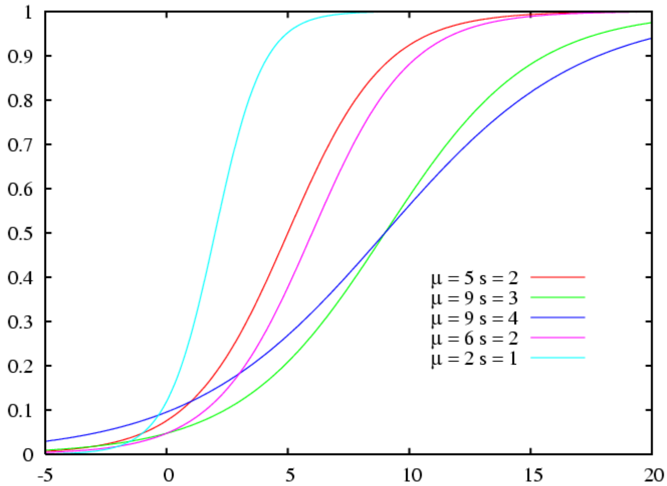


Ilustración 9 FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD LOGISTICA

Su función de densidad es:

$$f(x;\mu,s) = \frac{e^{-(x-\mu)/s}}{s\left(1+e^{-(x-\mu)/s}\right)^2}$$

$$= \frac{1}{4\,s}\;\operatorname{sech}^2\!\left(\frac{x-\mu}{2\,s}\right).$$

Nótese que puede expresarse en función del cuadrado de la secante hiperbólica "sech".

Si se realiza la sustitución $\sigma^2 = \pi^2\,s^2/3$, la función de densidad queda de la forma:

$$g(x;\mu,\sigma) = f(x;\mu,\sigma\sqrt{3}/\pi) = \frac{\pi}{\sigma\,4\sqrt{3}}\,\operatorname{sech}^2\!\left(\frac{\pi}{2\sqrt{3}}\,\frac{x-\mu}{\sigma}\right).$$

Teniendo una función de densidad de probabilidad definida por la siguiente gráfica:

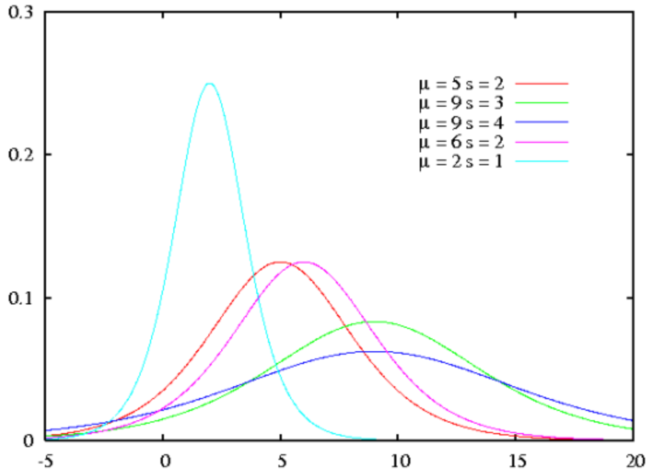


Ilustración 10 FUNCIÓN DE DENSIDAD DE PROBABILIDAD LOGISTICA

- <u>DISTRIBUCIÓN GAMMA:</u>

La distribución gama nos ayuda con problemas como el siguiente, supóngase que una pieza metálica se encuentra sometida a dicha fuerza, de manera que se romperá después de aplicar un número especifico de ciclos de fuerza. Si los ciclos ocurren de manera independientemente y una frecuencia promedio, entonces el tiempo de que debe transcurrir antes de que el material se rompa es una variable aleatoria que cumple con la distribución gama.

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

$$f(x;\alpha,\theta) = \begin{cases} \dfrac{x^{(\alpha-1)\exp\left(-\frac{\alpha}{\theta}\right)}}{\Gamma(\alpha)\theta^{\alpha}}, & x > 0;\ \alpha,\theta > 0, \\ 0, & Para\ cualquier\ otro\ caso \end{cases}$$

En donde $\Gamma(\alpha)$ es la función gama:

$$\Gamma(\alpha) = \int_0^{\infty} u^{(\alpha-1)\exp^{-u}}\, du$$

La distribución gama es muy versátil puesto que exhibe varios perfiles que dependen del valor del parámetro α. A continuación mostramos como queda representada la función de densidad, teniendo en cuenta que k=α.

Quedando la siguiente imagen como la gráfica de la función de densidad de probabilidad:

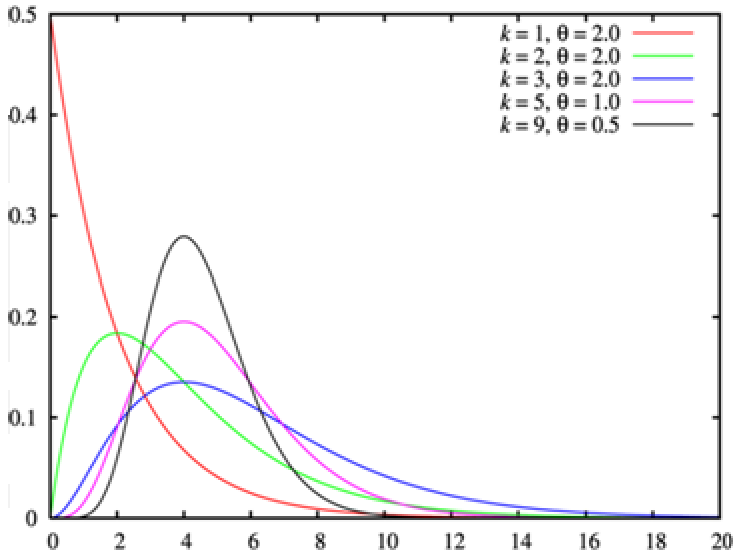


Ilustración 11 FUNCIÓN DE DENSIDAD DE PROBABILIDAD GAMMA

El valor esperado y la varianza de una variable aleatoria X de distribución gamma son

$$E[X] = k/\lambda = k\theta$$

$$V[X] = k/\lambda^2 = k\theta^2$$

La función de distribución acumulativa se determina por la siguiente función:

$$F(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} \int_0^x t^{(\alpha-1)} \exp\left(-\frac{t}{\theta}\right) dt, \qquad x > 0.$$

Función de distribución de probabilidad:

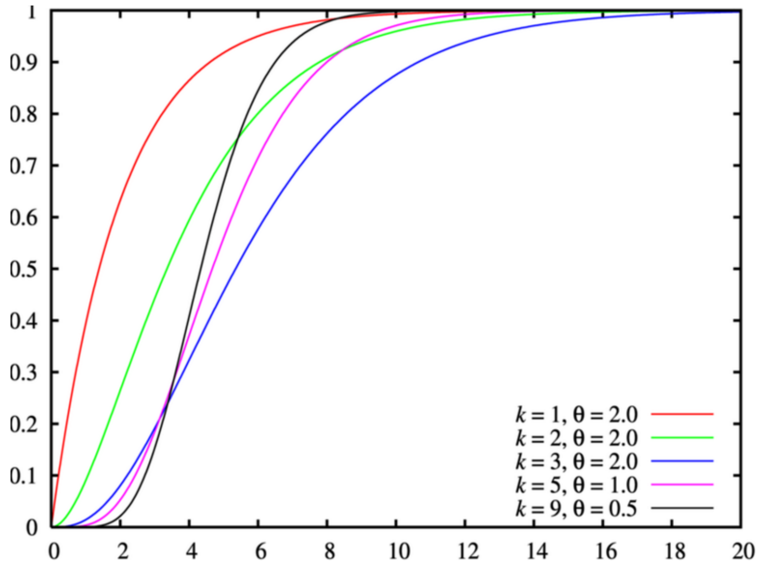


Ilustración 12 FUNCIÓN DE DENSIDAD DE PROBABILIDAD GAMMA

- <u>DISTRIBUCIÓN BETA:</u>

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

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

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)} x^{\alpha-1}(1-x)\text{^}(\beta-1), & 0 < x < 1,\ \ \alpha,\beta > 0, \\ 0, \quad para\ caulaquier\ otro\ valor \end{cases}$$

Las cantidades de α y β de la distribución beta son, ambas, parámetros de perfil. Valores distintos de α y β darán distintos perfiles para la función de densidad de beta. Si tanto como α como β son menores que uno, la distribución tiene forma de. Si α<1 y β≥ 1, la distribución tiene un perfil de J transpuesta, y si β<1 y α≥ 1, el perfil es una J. Cuando α y β son ambos mayores que uno, la distribución presenta un pico en $x = \frac{\alpha-1}{\alpha+\beta-2}$. Finalmente, la distribución beta es simétrica cuando α = β.

En la siguiente imagen mostramos el comportamiento de la función de densidad para diferentes tipos de α y β.

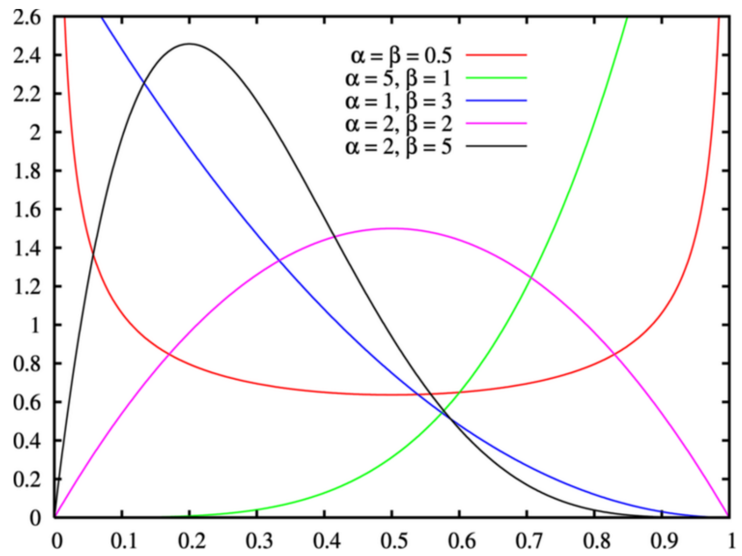


Ilustración 13 FUNCIÓN DE DENSIDAD DE PROBABILIDAD BETA

La función de distribución acumulativa se encuentra definida por:

$$P(X \leq x) = F(x, \alpha, \beta) = \begin{cases} 0 & x \leq 0, \\ \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)} \displaystyle\int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,dt & 0 < x < 1, \\ 1 & x \geq 1, \end{cases}$$

Quedando una gráfica de la siguiente manera:

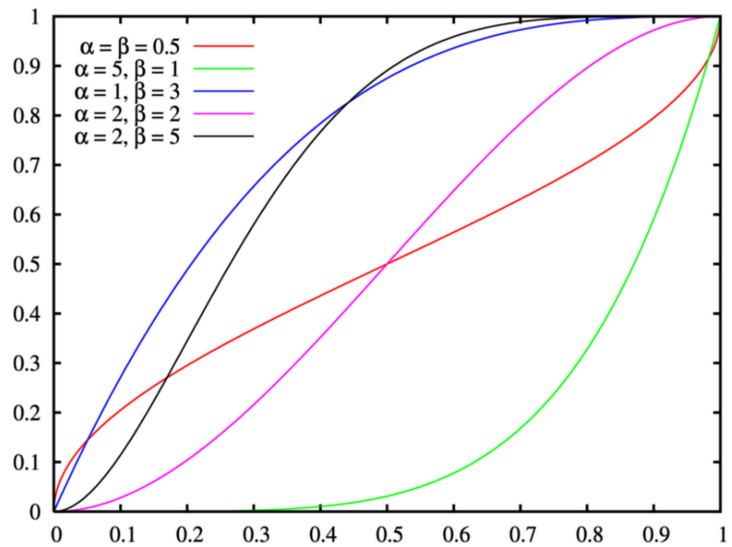


Ilustración 14 FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD BETA

- DISTRIBUCIÓN EXPONENCIAL:

En estadística la distribución exponencial es una distribución de probabilidad continua con un parámetro $\lambda > 0$ cuya función de densidad es:

$$f(x) = P(x) = \begin{cases} \lambda e^{-\lambda x} & \text{para } x \geq 0 \\ 0 & \text{en caso contrario} \end{cases}$$

Su función de distribución acumulada es:

Aquí mostramos la gráfica que dibuja la función de densidad de probabilidad:

En estadística la distribución exponencial es una distribución de probabilidad continua con un parámetro $\lambda > 0$ cuya función de densidad es:

$$f(x) = P(x) = \begin{cases} \lambda e^{-\lambda x} & \text{para } x \geq 0 \\ 0 & \text{en caso contrario} \end{cases}$$

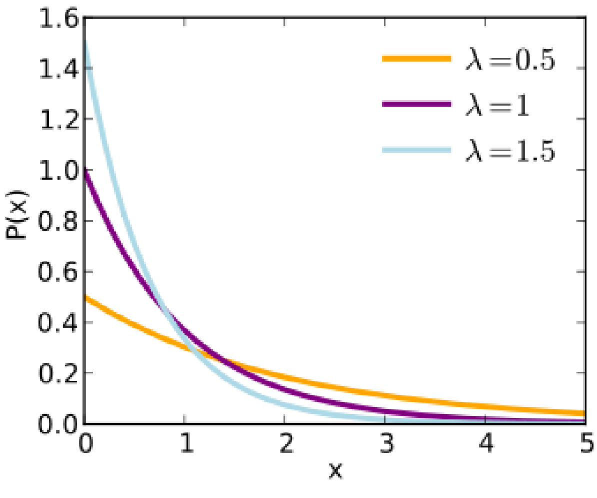


Ilustración 15 FUNCIÓN DE DENSIDAD DE PROBABILIDAD EXPONENCIAL

Su función de distribución acumulada es:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{para } x < 0 \\ 1 - e^{-\lambda x} & \text{para } x \geq 0 \end{cases}$$

Donde e representa el número e.

El valor esperado y la varianza de una variable aleatoria X con distribución exponencial son:

Y ahora mostramos la gráfica que describe la función de distribución de probabilidad acumulada:

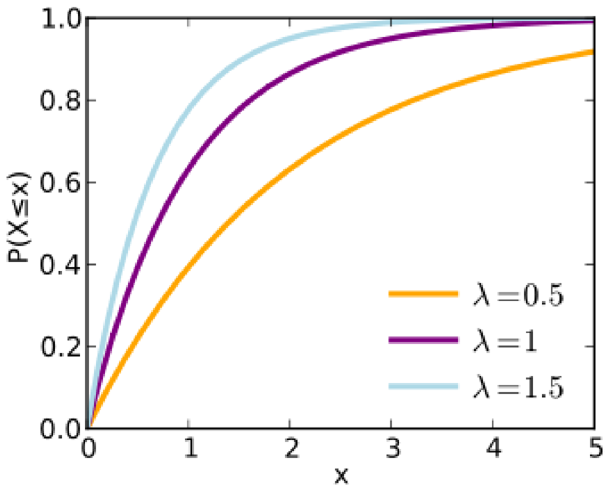


Ilustración 16 FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD EXPONENCIAL

$$E[X] = \frac{1}{\lambda}, \qquad V(X) = \frac{1}{\lambda^2}$$

La distribución exponencial es un caso particular de distribución gamma con k = 1. Además la suma de variables aleatorias que siguen una misma distribución exponencial es una variable aleatoria expresable en términos de la distribución gamma.

- DISTRIBUCIÓN F DE FISHER:

Usada en teoría de probabilidad y estadística, la distribución F es una distribución de probabilidad continua. También se le conoce como distribución F de Snedecor (por George Snedecor) o como distribución F de Fisher-Snedecor (porRonald Fisher).

Una variable aleatoria de distribución F se construye como el siguiente cociente:

$$F = \frac{U_1/d_1}{U_2/d_2}$$

Donde:

U1 y U2 siguen una distribución chi-cuadrado con d1 y d2 grados de libertad respectivamente, y U1 y U2 son estadísticamente independientes.

La distribución F aparece frecuentemente como la distribución nula de una prueba estadística, especialmente en el análisis de varianza. Véase el test F.

La función de densidad de una F(d1, d2) viene dada por

$$g(x) = \frac{1}{\mathrm{B}(d_1/2, d_2/2)} \left(\frac{d_1\,x}{d_1\,x + d_2}\right)^{d_1/2} \left(1 - \frac{d_1\,x}{d_1\,x + d_2}\right)^{d_2/2} x^{-1}$$

La función de densidad representa una gráfica tal que así:

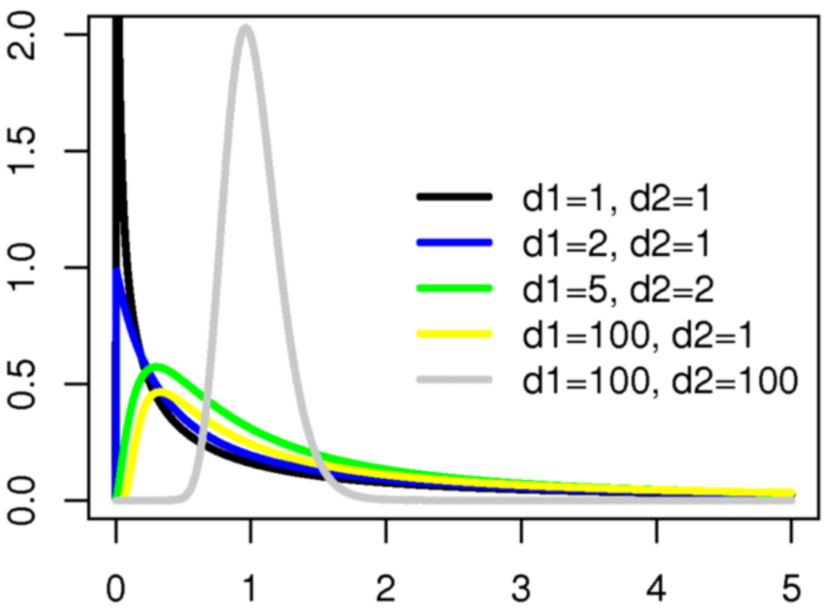


Ilustración 17 FUNCIÓN DE DENSIDAD DE PROBABILIDAD F DE FISHER

Para todo número real x ≥ 0, donde d1 y d2 son enteros positivos, y B es la función beta.

La función de distribución es

$$G(x) = I_{\frac{d_1 x}{d_1 x + d_2}}(d_1/2, d_2/2)$$

Y la función de distribución representa una gráfica tal que así:

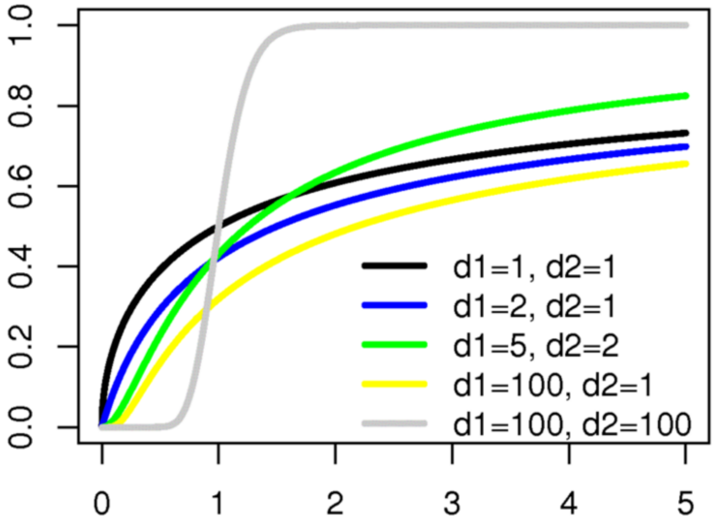


Ilustración 18 FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD F DE FISHER

Donde I es la función beta incompleta regularizada.

- <u>DISTRIBUCIÓN T DE STUDENT:</u>

En probabilidad y estadística, la distribución t (de Student) es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño.

Aparece de manera natural al realizar la prueba t de Student para la determinación de las diferencias entre dos medias muestrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones cuando se desconoce la desviación típica de una población y ésta debe ser estimada a partir de los datos de una muestra.

La distribución t de Student es la distribución de probabilidad del cociente

$$T = \frac{Z}{\sqrt{V/\nu}} = Z\sqrt{\frac{\nu}{V}}$$

La gráfica que caracteriza dicha distribución tiene una forma tal que así:

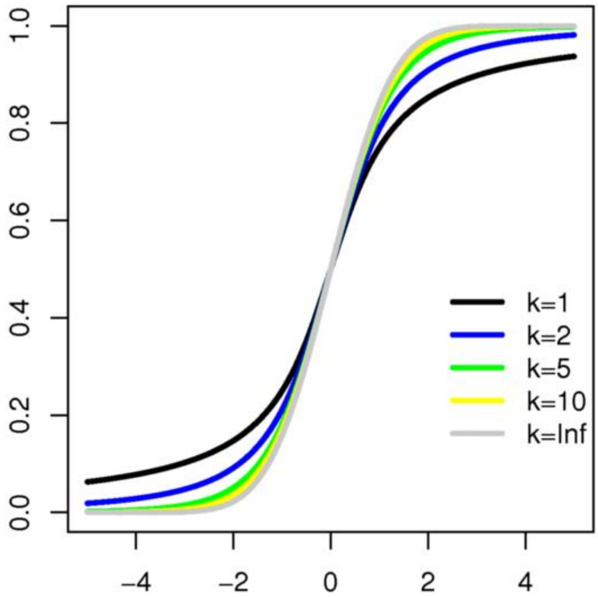


Ilustración 19 FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD T DE STUDENT

Donde

Z es una variable aleatoria distribuida según una normal típica (de media nula y varianza 1).

V es una variable aleatoria que sigue una distribución χ^2 con ν grados de libertad.

Z y V son independientes

Si μ es una constante no nula, el cociente $\frac{Z+\mu}{\sqrt{V/\nu}}$ es una variable aleatoria que sigue la distribución t de Student no central con parámetro de no-centralidad μ.

Aparición y especificaciones de la distribución t de Student[]

Supongamos que X1,..., Xn son variables aleatorias independientes distribuidas normalmente, con media μ y varianzaσ2. Sea

$$\overline{X}_n = (X_1 + \cdots + X_n)/n$$

La media muestral. Entonces

$$Z = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

Sigue una distribución normal de media 0 y varianza 1.

Sin embargo, dado que la desviación estándar no siempre es conocida de antemano, Gosset estudió un cociente relacionado,

$$T = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}},$$

$$S^2(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Es la varianza muestral y demostró que la función de densidad de T es

$$f(t) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}(1 + t^2/\nu)^{-(\nu+1)/2}$$

Teniendo una gráfica así:

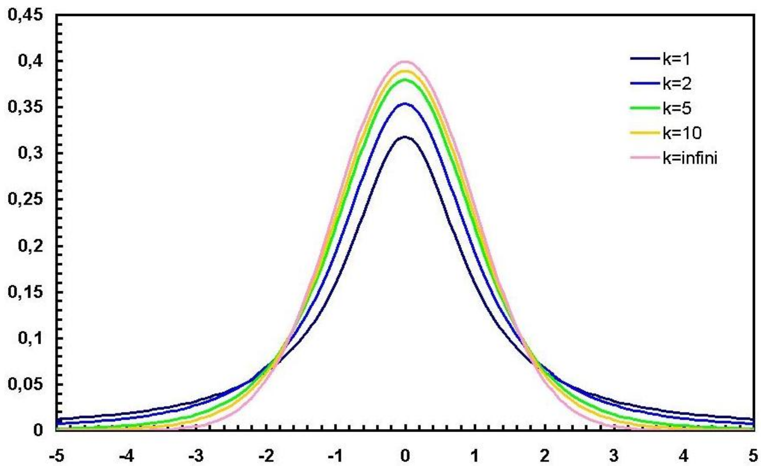


Ilustración 20 FUNCIÓN DE DENSIDAD DE PROBABILIDAD T DE STUDENT

Donde ν es igual a n − 1.

La distribución de T se llama ahora la distribución-t de Student.

El parámetro ν representa el número de grados de libertad. La distribución depende de ν, pero no de μ o σ, lo cual es muy importante en la práctica.

El procedimiento para el cálculo del intervalo de confianza basado en la t de Student consiste en estimar la desviación típica de los datos S y calcular el error estándar de la media: $\overline{X} = \frac{S}{\sqrt{n}}$, siendo entonces el intervalo de confianza para la media: $\overline{X} = \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$.

Es este resultado el que se utiliza en el test de Student: puesto que la diferencia de las medias de muestras de dos distribuciones normales se distribuye también normalmente, la distribución t puede usarse para examinar si esa diferencia puede razonablemente suponerse igual a cero.

Para efectos prácticos el valor esperado y la varianza son:

$E(t(n)) = 0$ Y $Var(t(n-1)) = n/(n-2)$ para $n > 3$ $E(t(n)) = 0$ y $Var(t(n-1)) = n/(n-2)$ para $n > 3$ $E(t(n)) = 0$ y $Var(t(n-1)) = n/(n-2)$ para $n > 3$ $E(t(n)) = 0$ y $Var(t(n-1)) = n/(n-2)$ para $n > 3$

- ## DISTRIBUCIÓN DE CAUCHY:

La distribución Cauchy-Lorentz, llamada en honor a Augustin Cauchy y Hendrik Lorentz, es una distribución de probabilidad continua. Es conocida como la distribución de Cauchy y en el ámbito de la física se conoce como la distribución de Lorentz, la función Lorentziana ó la distribución de Breit-Wigner. Su importancia en la física es dada por ser la solución de la ecuación diferencial que describe la resonancia forzada. En espectroscopia describe la forma de las líneas espectrales que son ampliadas por diversos mecanismos, en particular, el mecanismo de ensanchamiento por colisión.

En estadística la distribución de Cauchy (a veces también distribución de Lorentz) es una probabilidad continua cuya función de densidad es

$$f(x;x_0,\gamma)=\frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}$$
$$=\frac{1}{\pi}\left[\frac{\gamma}{(x-x_0)^2+\gamma^2}\right]$$

Donde x0 es el parámetro de corrimiento que especifica la ubicación del pico de la distribución, y γ es el parámetro de escala que especifica el ancho medio al máximo medio (half-width at half-maximum, HWHM).

En el caso especial donde x0 = 0 y γ = 1 es denominado la distribución estándar Cauchy con la función de densidad de probabilidad

$$f(x;0,1)=\frac{1}{\pi(1+x^2)}.$$

Quedando representada de las siguientes maneras:

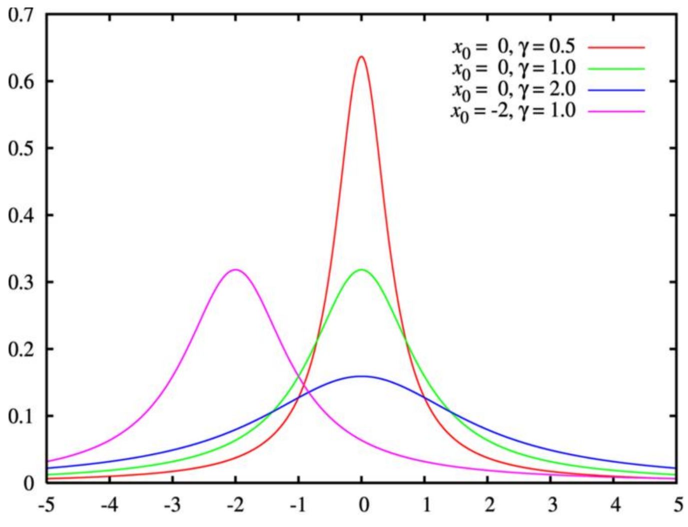


Ilustración 21 FUNCIÓN DE DENSIDAD DE PROBABILIDAD CAUCHY

En general la distribución de Cauchy no tiene valor esperado ni varianza.

Sean U y V dos variables aleatorias uniformes dentro -1 y 1 y $U^2+V^2<1$, entonces el número U/V tiene la distribución Cauchy.

La función de distribución acumulativa (CDF) es:

$$F(x;x_0,\gamma)=\frac{1}{\pi}\arctan\left(\frac{x-x_0}{\gamma}\right)+\frac{1}{2}$$

y la función inversa de distribución acumulativa para la distribución Cauchy es

$$F^{-1}(p;x_0,\gamma)=x_0+\gamma\,\tan\left[\pi\,\left(p-\tfrac{1}{2}\right)\right].$$

Como les mostramos en la siguiente imagen:

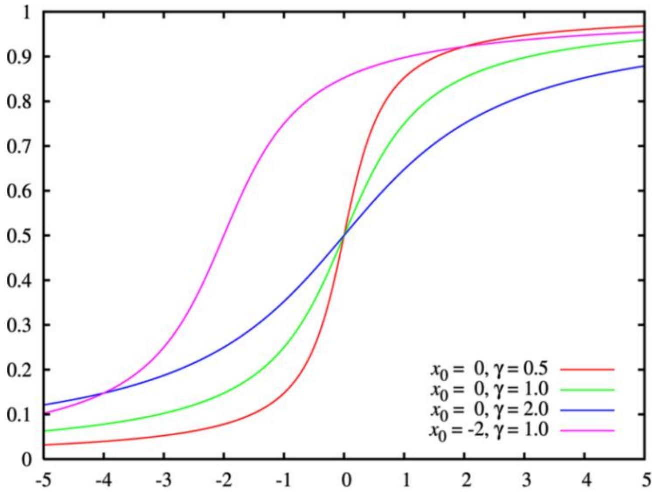


Ilustración 22 FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD CAUCHY

- <u>DISTRIBUCIÓN CHI-CUADRADO:</u>

En estadística, la distribución de Pearson, llamada también ji cuadrada(o) o chi cuadrado(a) (χ^2), es una distribución de probabilidad continua con un parámetro k que representa los grados de libertad de la variable aleatoria

$$X=Z_1^2+\cdots+Z_k^2$$

Donde Z_i son variables aleatorias normales independientes de media cero y varianza uno. El que la variable aleatoria X tenga ésta distribución se representa habitualmente así: $X\sim\chi^2_k$.

Su función de densidad es:

$$f(x;k) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} & \text{para } x > 0, \\ 0 & \text{para } x \leq 0 \end{cases}$$

Donde Γ es la función gamma.

Aquí se ve reflejada dicha función:

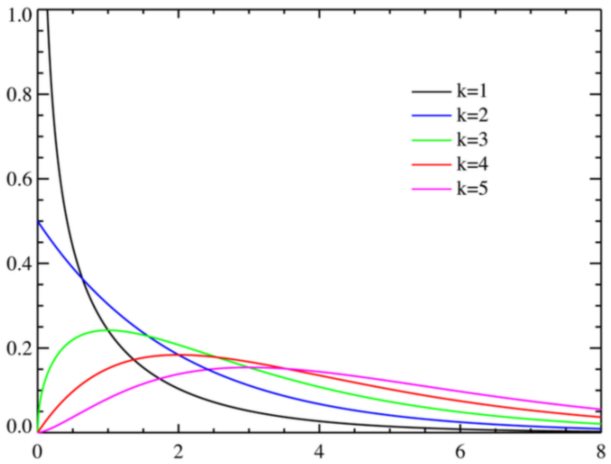


Ilustración 23 FUNCIÓN DE DENSIDAD DE PROBABILIDAD CHI-CUADRADO

Su función de distribución es

$$F_k(x) = \frac{\gamma(k/2, x/2)}{\Gamma(k/2)}$$

Donde $\gamma(k, z)$ es la función gamma incompleta.

El valor esperado y la varianza de una variable aleatoria X con distribución χ^2 son, respectivamente, k y 2k.

Quedando la siguiente gráfica como resultad de la función de distribución de probabilidad:

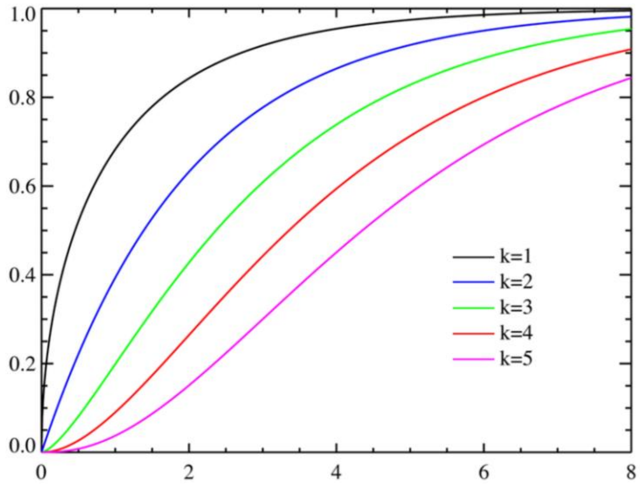


Ilustración 24 FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD CHI-CUADRADO

La distribución χ² tiene muchas aplicaciones en inferencia estadística. La más conocida es la de la denominada prueba χ² utilizada como prueba de independencia y como prueba de bondad de ajuste y en la estimación de varianzas. Pero también está involucrada en el problema de estimar la media de una población normalmente distribuida y en el problema de estimar la pendiente de una recta de regresión lineal, a través de su papel en la distribución t de Student.

Aparece también en todos los problemas de análisis de varianza por su relación con la distribución F de Snedecor, que es la distribución del cociente de dos variables aleatorias independientes con distribución χ².

4.2.4. DEMANDA ALEATORIA

Para saber qué tipo de distribución de probabilidad sigue alguna serie de datos tenemos que realizar sobre estos un análisis, el cual nos determinará su distribución.

Los costes dentro de este tipo de demanda son los mismos que para el caso determinista. El costo general del lote óptimo está directamente relacionado con el número de unidades demandadas.

Uno de los grandes retos de este trabajo fue como establecer una lectura de los datos introducidos por los clientes, determinando así su distribución, como hemos visto anterior mente en el apartado [4] explicamos una gran variedad de funciones que determinan una probabilidad aleatoria.

Con esto lo que queremos decir es que la clave del trabajo es determinar un tipo de distribución teniendo en cuenta los datos de una serie.

4.2.5. AJUSTE DE LA DEMANDA

Para establecer un ajuste de la demanda a una función tenemos que establecer una serie de parámetros que nos sirvan de indicador.

Estudiaremos los datos teniendo en cuenta que lo que se busca es una aproximación óptima para determinar el tipo de distribución asociada a una entrada de datos.

- AJUSTE POR MINIMIZACIÓN DE ERRORES:

El ajuste por minimización de errores, también conocido como *Estimación por mínimos cuadrados para el modelo lineal simple.*

En esta sección se estudiará la estimación por mínimos cuadrados para el modelo lineal simple en el que solo tienen una variable de predicción (como es nuestro caso), y se supone una ecuación de regresión lineal.

Todo tipo de regresión lineal se determinará por la siguiente función:

$$Yi = \beta o + \beta 1xi + \varepsilon i \qquad i = 1,2,3 \dots n,$$

Donde Yi es la i-esima observación de la variable respuesta, la cual corresponde al i-esimo valor xi de la variable de predicción, εi es el error aleatorio no observable asociado con Yi ; βo y $\beta 1$ son los parámetros desconocidos que representan la intersección y la pendiente, respectivamente. Dicha expresión se conoce como modelo lineal simple.

Cada observación Yi es una variable aleatoria que es la suma de dos componentes; eltermino no aleatorio $\beta o + \beta 1xi$, y la componente aleatoria εi. Si εi fuera un valor igual a cero, la observación Yi se encontraría sobre la línea de regresión $\beta o + \beta 1xi$. Por lo tanto, εi es la distancia vertical de la observación a la línea de regresión. Dado que se supone:

$$E(\varepsilon i) = 0, \quad Var(\varepsilon i) = \sigma^2 \quad i = 1,2, \ldots., n,$$

$$Cov(\varepsilon i, \varepsilon j) = 0, \quad i \neq j;$$

Despejando el erro nos queda la ecuación principal

$$\varepsilon i = Yi - (\beta 0 + \beta 1 xi)$$

Aquí aplicamos la suma de los cuadrados.

$$\sum_{i=1}^{n} \varepsilon i\text{^}2 = \sum_{i=1}^{n} (Yi - (\beta 0 + \beta 1 xi))\text{^}2$$

De aquí obtenemos b0 y b1.

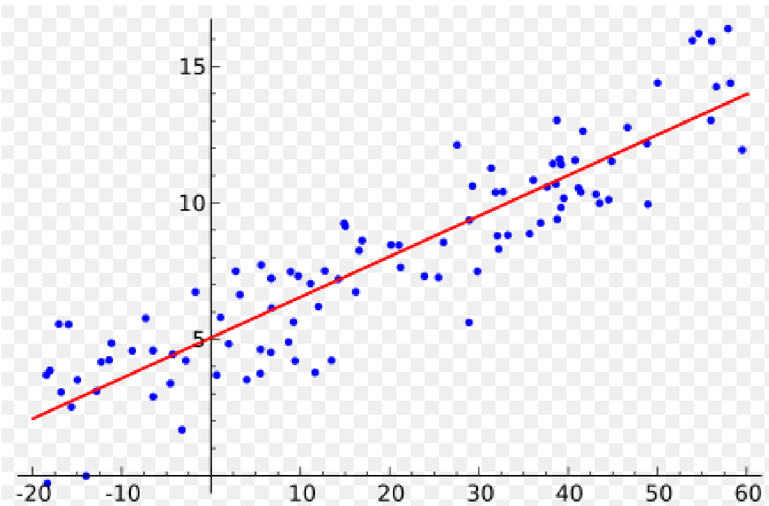
$$b1 = (\frac{\sum_{i=1}^{n}(xi - \chi)(yi - \gamma))}{\sum_{i=1}^{n}(xi - \chi)^2})$$

$$b0 = \gamma - (\frac{\sum_{i=1}^{n}(xi - \chi)(yi - \gamma))}{\sum_{i=1}^{n}(xi - \chi)^2})\, \chi$$

Siendo la siguiente una gráficas de regresión lineal:

La recta de esta gráfica está definida por:

$$\ddot{y}i = bo + b1xi$$



I.AJUSTE POR MINIMIZACIÓN DE ERRORES

$\ddot{y}i$: Es la distancia entre un punto xi y al recta medido en el eje vertical.

Como podemos concluir este modelo matemático nos da el valor de las diferencias entre un punto y su línea de regresión, lo cual nos ayudará a calcular el valor del error teniendo en cuenta los tipos de distribuciones. Cabe destacar que este método no contribuye a determinar que pasen sucesos futuros en dicha línea de tendencia, pero que la distribución tendrá que estar próxima a esta.

- CRITERIO DE INFORMACIÓN DE AKAIKE (AIC):

El criterio de información de Akaike (AIC) es una medida de la calidad relativa de un modelo estadístico, para un conjunto dado de datos. Como tal, el AIC proporciona un medio para la selección del modelo.

AIC maneja un trade-off entre la bondad de ajuste del modelo y la complejidad del modelo. Se basa en la entropía de información: se ofrece una estimación relativa de la información pérdida cuando se utiliza un modelo determinado para representar el proceso que genera los datos.

AIC no proporciona una prueba de un modelo en el sentido de probar una hipótesis nula, es decir AIC puede decir nada acerca de la calidad del modelo en un sentido absoluto. Si todos los modelos candidatos encajan mal, AIC no dará ningún aviso de ello.

En el caso general, la AIC es

$$AIC = 2k - 2\ln(L)$$

Donde k es el número de parámetros en el modelo estadístico, y L es el máximo valor de la función de verosimilitud para el modelo estimado.

Dado un conjunto de modelos candidatos para los datos, el modelo preferido es el que tiene el valor mínimo en el AIC. Por lo tanto AIC no sólo recompensa la bondad de ajuste, sino también incluye una penalidad, que es una función creciente del número de parámetros estimados. Esta penalización desalienta el sobreajuste (aumentando el número de parámetros libres en el modelo mejora la bondad del ajuste, sin importar el número de parámetros libres en el proceso de generación de datos).

AIC se basa en la teoría de la información. Supongamos que los datos se generan por algún proceso desconocido f. Consideremos dos modelos candidatos para representar f: g1 y g2. Si supiéramos f, entonces podríamos encontrar la información perdida del uso de g1 para representar f calculando la divergencia de Kullback-Leibler , D KL (f ‖ g 1), de manera similar, la información perdida del uso de g2 para representar f sería obtenido calculando D KL (f ‖ 2 g). Entonces nos volveríamos a elegir el modelo candidato que minimiza la pérdida de información.

- Función de verosimilitud

En estadística, la función de verosimilitud (o, simplemente, verosimilitud) es una función de los parámetros de un modelo estadístico que permite realizar inferencias acerca de su valor a partir de un conjunto de observaciones.

No debe confundirse con el término probabilidad: ésta permite, a partir de una serie de parámetros conocidos, realizar predicciones acerca de los valores que toma una variable aleatoria.

En cierto sentido, la verosimilitud es una versión inversa de la probabilidad condicional. Conocido un parámetro B, la probabilidad condicional de A es P(A|B), pero si se conoce A, pueden realizarse inferencias sobre el valor de B gracias al teorema de Bayes, según el cual

$$P(B \mid A) = \frac{P(A \mid B)\, P(B)}{P(A)}.$$

La función de verosimilitud, L(b |A), definida como

$$L(b \mid A) = P(A \mid B = b),$$

Desempeña el mismo papel bajo un enfoque no bayesiano. De hecho, lo relevante no es el valor en sí de L(b |A) sino la razón de verosimilitudes,

$$\frac{L(b_2|A)}{L(b_1|A)},$$

Que permite comparar cuanto más verosímil es el parámetro b_1 que el b_2 a la hora de explicar el evento A. De ahí que en ocasiones se entienda que la función de verosimilitud, más que una función en sí, sea la clase de funciones

$$L(b \mid A) = \alpha \; P(A \mid B = b),$$

Donde α es una constante de proporcionalidad.

La función de verosimilitud, abundando en los razonamientos anteriores, abre la vía para dos técnicas muy habituales en inferencia estadística: las de la máxima verosimilitud y la del test de la razón de verosimilitudes.

- ## CRITERIO DE INFORMACIÓN BAYESIANA BIC:

En estadística, el criterio de información bayesiano (BIC) o el más general criterio de Schwarz (SBC también, SBIC) es un criterio para la selección de modelos entre un conjunto finito de modelos. Se basa, en parte, de la función de probabilidad y que está estrechamente relacionado con el Criterio de Información de Akaike (AIC).

Cuando el ajuste de modelos, es posible aumentar la probabilidad mediante la adición de parámetros, pero si lo hace puede resultar en sobreajuste. Tanto el BIC y AIC resuelven este problema mediante la introducción de un término de penalización para el número de parámetros en el modelo, el término de penalización es mayor en el BIC que en el AIC.

El BIC fue desarrollado por Gideon E. Schwarz, quien dio un argumento bayesiano a favor de su adopción.1 Akaike también desarrolló su propio formalismo Bayesiano, que ahora se conoce como la ABIC por Criterio de Información Bayesiano de Akaike".2

Matemáticamente hablando:

El BIC es una consecuencia derivada asintótica bajo los supuestos de que la distribución de los datos se encuentra en la familia exponencial. Vamos:

- x = los datos observados;
- n = el número de datos u observaciones x, o equivalentemente, el tamaño de la muestra;

- k = el número de parámetros libres a ser estimados. Si el modelo está bajo el supuesto de que es lineal, k es el número de regressores, incluyendo el intercepto;
- $p(x|M)$ = La probabilidad marginal de los datos observados dado el modelo M; esto es, Es decir, la integral de la función de verosimilitud $p(x|\theta, M)$ veces la distribución de probabilidad antes $p(\theta|M)$ sobre los parámetros θ del modelo M para los datos observados fijos x ;
- $\hat{L}$ = El máximo valor de la función de verosimilitud del modelo M, i.e. $\hat{L} = p(x|\hat{\theta}, M)$, donde $\hat{\theta}$ son los valores de los parámetros que maximizan la función de verosimilitud.

La fórmula para el BIC es:

$$-2 \cdot \ln p(x|M) \approx \mathrm{BIC} = -2 \cdot \ln \hat{L} + k \ln(n).$$

Bajo la suposición de que los errores de modelo o perturbaciones son independientes e idénticamente distribuidos según una distribución normal y que la condición límite de que la derivada de la probabilidad de registro con respecto a la varianza real es cero, esto se convierte en (hasta una constante aditiva, la cual sólo depende de n, y no en el modelo): [3]

$$\mathrm{BIC} = n \cdot \ln(\widehat{\sigma_e^2}) + k \cdot \ln(n)$$

Donde $\widehat{\sigma_e^2}$ es la varianza del error.

La varianza del error, en este caso se define como:

$$\widehat{\sigma_e^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2.$$

Uno puede señalar de teoría de la probabilidad de que $\widehat{\sigma_e^2}$ es un estimador sesgado de la varianza verdadera, σ^2. Dejar $\widehat{\widehat{\sigma_e^2}}$ denotar la forma no sesgada de la aproximación de la varianza del error. Se define como:

$$\widehat{\widehat{\sigma_e^2}} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2.$$

Adicionalmente, bajo el supuesto de normalidad la siguiente versión puede ser más manejable:

$$\mathrm{BIC} = \chi^2 + k \cdot \ln(n).$$

Tenga en cuenta que hay una constante añadido que se deriva de transición de la log-verosimilitud para χ^2, sin embargo, en el uso del BIC para determinar el "mejor" modelo de la constante se convierte en trivial.

Dadas dos modelos estimados, el modelo con el menor valor de BIC es el que se prefiere. El BIC es un aumento de la función de σ_e^2 y una función creciente de k. Es decir, la variación no explicada en la variable dependiente y el número de variables explicativas aumentan el valor de BIC. Por lo tanto, menor BIC implica un número menor de variables explicativas, mejor ajuste, o ambos. La fuerza de la evidencia en contra del modelo con el mayor valor de BIC se puede resumir de la siguiente manera:

- PRUEBA DE KOLMOGOROV-SMIRNOV:

Dicha prueba no necesita una muestra grande y tampoco que los datos estén agrupados. Esta se basa en la comparación entre las funciones de distribución acumulativa que se observan en la muestra ordenada y la distribución propuesta bajo una hipótesis nula de que la distribución $Fo(x)$, se rechaza.

Función acumulativa:

$$Sn(x) = \begin{cases} 0 & x < x(1), \\ \frac{k}{n} & x(k) \leq x < x(k+1), \\ 1 & x \geq x(n) \end{cases}$$

Donde $x(1), x(2), \ldots.., x(n)$, son las observaciones de la muestra de un tamaño n.

Para cualquier valor de x de la muestra aleatoria, $Sn(x)$ es la proporción del número de valores en la muestra que son iguales o menores a x. Ya que $Fo(x)$ se encuentra completamente especificada, es posible evaluar a $Fo(x)$ para algún valor deseado de x, entonces comparar ese último con el valor correspondiente de $Sn(x)$. Si la hipótesis nula es verdadera, entonces es lógico esperar que la diferencia sea realmente pequeña. La estadística de Kolmogorov-Smirnov se define como.

$$D(n) = máx|Sn(x) - Fo(x)|$$

4.3. PORQUE RSTUDIO:

Como hemos podido observar dicha herramienta de trabajo la conocía antes de llegar al proyecto de fin de grado, y una vez establecido los objetivos antes mencionados y la parte teórica del Hito I, observamos que la herramienta tiene que cumplir con todo lo anterior.

- ✓ Realiza cálculos estadísticos determinado las probabilidades.
- ✓ Tengo experiencia previa con este programa debido a su similitud con "R".
- ✓ Podemos usar paquetes que nos facilitan la creación de interfaces agradables.
- ✓ Es fácil y como de usar.
- ✓ Nos ofrece crear una interfaz en la que podemos introducir los objetivos anteriores como entrada.
- ✓ Dentro de la web de R-studio disponemos de una gran variedad de tutoriales explicando desde cero el funcionamiento de dicho programa.
- ✓ Es un software libre.
- ✓ Dispone de foros para establecer tus dudas.

Este conjunto de requisitos que nos brinda R-studio nos hace establecer este programa como determinante en el trabajo.

4.3.1. QUE ES RSTUDIO:

Comenzaremos con una breve descripción del programa:

R es un entorno y lenguaje de programación con un enfoque al análisis estadístico.

R es una implementación de software libre del lenguaje S pero con soporte de alcance estático. Se trata de uno de los lenguajes más utilizados en investigación por la comunidad estadística, siendo además muy popular en el campo de la minería de datos, la investigación biomédica, la bioinformática y las matemáticas financieras. A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo o graficación.

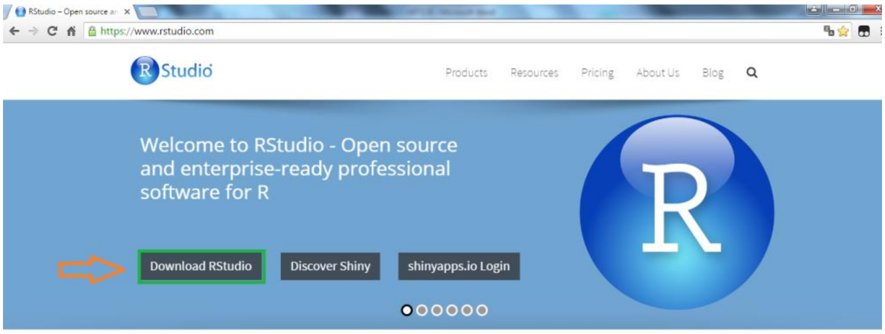
R es parte del sistema GNU y se distribuye bajo la licencia GNU GPL. Está disponible para los sistemas operativos Windows,Macintosh, Unix y GNU/Linux.

Dentro de lo que es el programa que vamos a usar no daremos una descripción detallada del programa ya que daría para otro trabajo de fin de grado, y como hemos establecido en el apartado anterior las ventajas son obvias.

Una vez explicadas las nociones básicas de R-studio procederemos a una mostrar cómo se puede instalar en nuestro ordenador.

- ## INSTALACIÓN R-STUDIO:

En primer lugar nos dirigiremos a la siguiente dirección web https://www.rstudio.com/ , como vemos en la siguiente imagen clicamos en el botón de descargar.



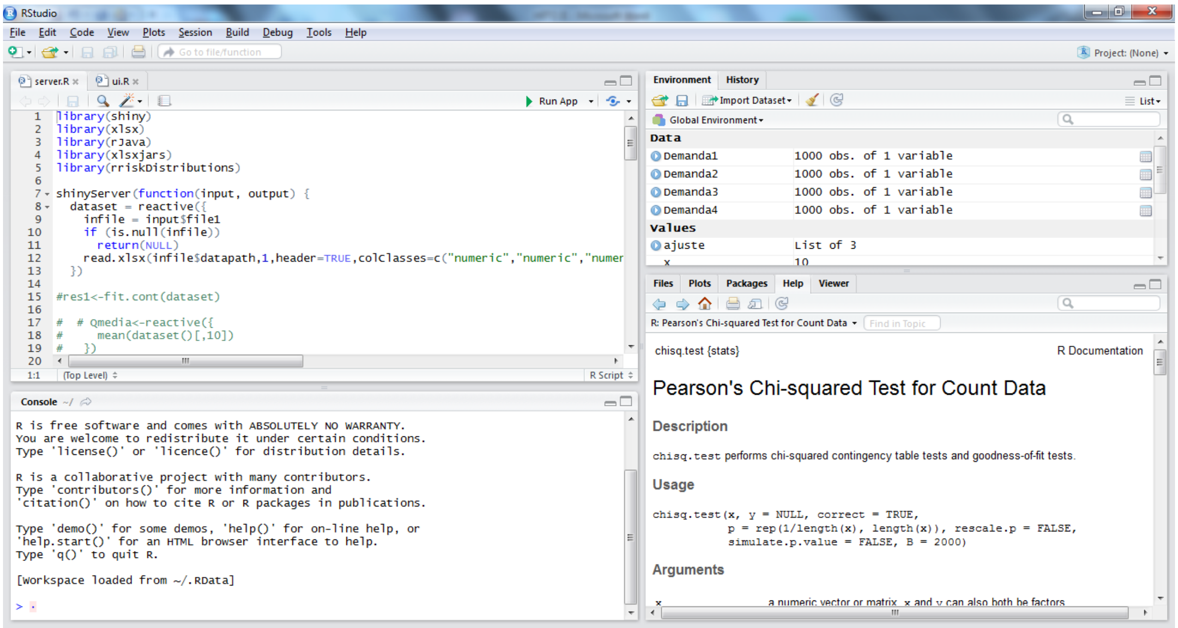
Una vez dentro de la descarga nos aparece la opción de quererlo instalar en la nube o en el escritorio, seleccionamos la que más nos convenga, también tiene la opción de compra de licencia para las empresas.

Una vez establecido lo anterior non mostrara una interfaz en la que aparecerán todas las opciones de descarga para las diferentes plataformas existentes, aquí deberá seleccionarla acorde con su ordenador.

Una vez descargado el ejecutable lo iniciamos

- ## FUNCIONAMIENTO BÁSICO:

En la siguiente imagen veremos cómo queda la interfaz del programa:



Como podemos observar en el margen superior izquierdo hay abierto dos archivos, uno es el 'server' el otro el 'ui'. En el archivo 'server' introducimos las funciones y algoritmos que usamos así como también los paquetes que usaremos. El otro archivo 'ui' es el que nos definirá la interfaz del programa.

Encima del cuadro inferior derecho encontraos un pestaña llamada paquetes, aquí es donde deberemos cargar los paquetes que vayamos a usar.

El programa esta creado de forma bastante intuitiva con cuadros e indicaciones de donde debemos buscar las cosas que precisemos, así como también podemos encontrar un cuadro de ayuda colocado cerca del de la instalación de los paquetes.

4.4. PAQUETES QUE HEMOS USADO:

Antes de empezar a explicar los tipos de paquetes que hemos utilizado explicaremos que son y cómo funcionan.

4.4.1. ¿Qué es un paquete?

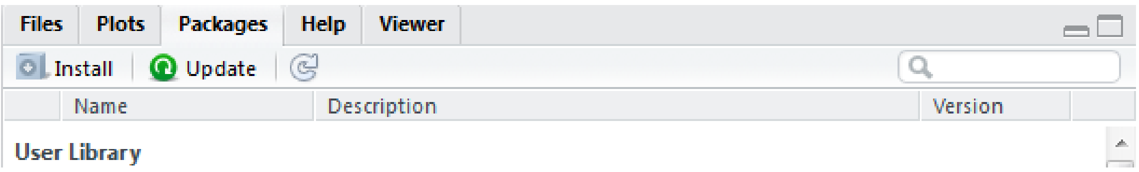
Dentro de R-studio esta la posibilidad de cargar paquetes, estos son programas creados anteriormente que tú puedes usar sin tener que escribir todo el código de nuevo en tu programa.

Como vemos este tipo de archivos nos dan una gran facilidad a la hora de hacer que nuestro trabajo de programación sea más rápido.

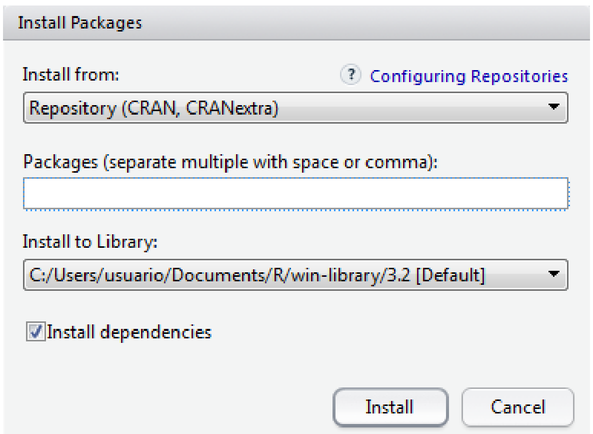
4.4.2. Paquetes:

- <u>INSTALACIÓN:</u>

Como vemos en la siguiente imagen, primero seleccionamos en la zona superior del cuadro inferior derecho el cuadro donde pone paquetes, y seleccionamos instalar para instalarlo en nuestro ordenador.



Una vez seleccionado nos aparece le siguiente cuadro de dialogo:



En este cuadro buscamos los paquetes que previamente hemos determinado que nos son necesarios y los instalamos.

Cuando ya tengamos instalados los paquetes en la librería solo deberemos nombrarlos desde el archivo 'server' para cargarlos como se ve en la siguiente imagen:

Los paquetes que hemos usado son los siguientes:

1. library(shiny)

   Este paquete es el responsable de crear una app, que bien determinaremos si la queremos en la nube o instalaremos en nuestro ordenador, en nuestro caso lo haremos desde nuestro ordenador.

2. library(xlsx)
3. library(rJava)
4. library(xlsxjars)

   Estos tres paquetes anteriormente nombrados son los encargados de establecer la lectura de un excel al introducirlo en el programa.

5. library(rriskDistributions)

   Este paquete es el encargado en ayudarnos a seleccionar que tipo de distribución llevan los datos cargados en el programa.

- <u>FUNCIONAMIENTO DE LOS PAQUETES:</u>

1. Paquete Shiny: este paquete como hemos determinado antes es el encargado de crear una app. El funcionamiento del mismo es bastante sencillo teniendo en cuenta todas las opciones que nos muestra, dentro de R-stuido Shyni encontramos tutoriales foros y ejemplos del funcionamiento.

En esta web encontraremos toda la ayuda necesaria para el manejo de dicho programa, http://shiny.rstudio.com/

2. Paquetes (xlsx) (rJava) (xlsxjars): estos tres paquetes nos ayudan a la lectura de un excel desde el mismo programa.

3. Paquete (rriskDistributions): este paquete hace referencia a lo nombrado desde la determinación del tipo de distribución hasta el tipo de ajuste mencionado en el Hito I, por lo que es un paquete que nos facilita bastante el funcionamiento del programa.

   Este hace una lectura de los datos y establece unos métodos de ajuste teniendo en cuenta los nombrados anteriormente.

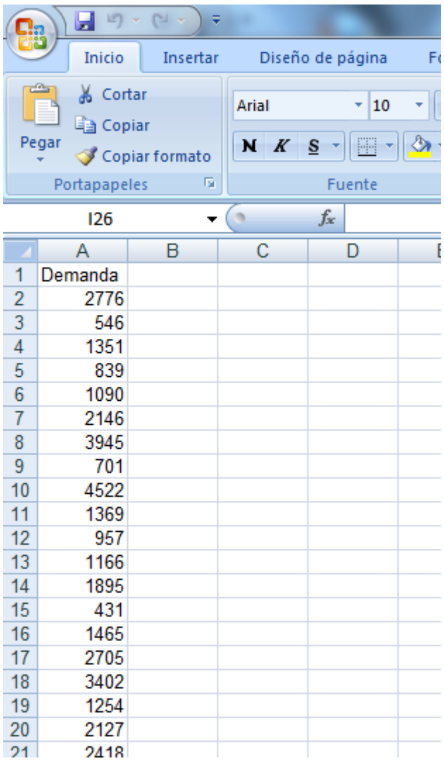
4.5. MODELO EXCEL:

Como hemos observado la entrada de datos parte de una tabla excel, habiendo introducido los paquetes anteriormente nombrados solo conseguimos su lectura, debido a que el programa le por columnas o filas como determinemos, tuvimos que establecer un modelo de excel para evitar problema. Este modelo no consiste más que en determinar si los datos están en filas o columnas y determinar cuál es.

En nuestro lugar decidimos establecer la lectura de columnas y establecer el número de esta, en nuestro caso siempre leerá la última columna y dejara la fila numero 1 disponible para títulos.

Con esto queremos decir que el programa leerá cualquier archivo excel en el cual la segunda columna este compuesta por datos numéricos, esto último es muy importante a la hora de crear un archivo, establecer en dicha columna el formato de datos numéricos.

Como vemos en la siguiente imagen:



	A	B	C	D
1	Demanda			
2	2776			
3	546			
4	1351			
5	839			
6	1090			
7	2146			
8	3945			
9	701			
10	4522			
11	1369			
12	957			
13	1166			
14	1895			
15	431			
16	1465			
17	2705			
18	3402			
19	1254			
20	2127			
21	2418			

4.6. FUNCIONAMIENTO DEL PROGRAMA:

Antes de comenzar debemos establecer las entradas y salidas de datos, que se podría resumir en que es lo que queremos y que es los que necesitamos.

- ✓ Queremos un programa que nos determine el lote óptimo.
- ✓ Necesitamos saber que costes de mantenimiento precisamos.
- ✓ Necesitamos saber que costes por faltantes tenemos.
- ✓ Necesitamos saber que costes por producir tenemos.
- ✓ Necesitamos saber que tipo de distribución tenemos.
- ✓ Necesitamos saber los consumos dentro de los intervalos determinados por los proveedores.
- ✓ Necesitamos saber el nivel de inventario inicial.
- ✓ Necesitamos saber el stock de seguridad.

Como podemos observar hay bastantes datos de entrada.

Antes de seguir decidimos asignar un nombre a dicho programa, para su distinción y facilitar su búsqueda en cualquier caso. El nombre será OPTINV, como contracción de optimización de inventarios.

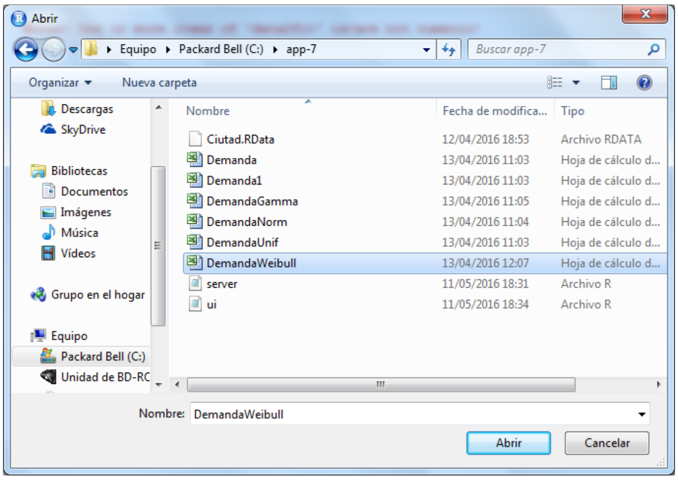
Una vez tengamos instalado el programa procederemos a ver su funcionamiento. En primera instancia la interfaz principal con la que nos encontraremos tendrá una apariencia así:

Como vemos tenemos las entradas de datos, y hemos creado una pestaña principal en la que se nos especifica la instrucción de uso, que explicaremos posteriormente detalladamente.

4.6.1. Ajuste de la demanda:

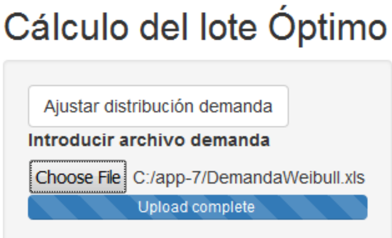
Para comenzar al ajuste de la demanda el primer paso será introducir el archivo excel, para ello clicamos en el botón “Choose file”, posteriormente seleccionamos el archivo que queramos analizar para ello este archivo debe cumplir los requisitos nombrados anteriormente.

En nuestro caso hemos elegido el siguiente archivo excel:



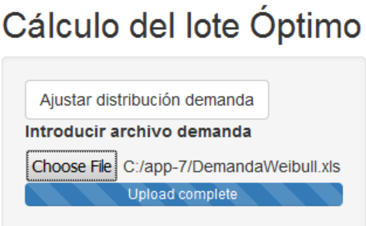
Clicamos en el botón de abrir.

Como vemos el archivo se ha cargado correctamente.

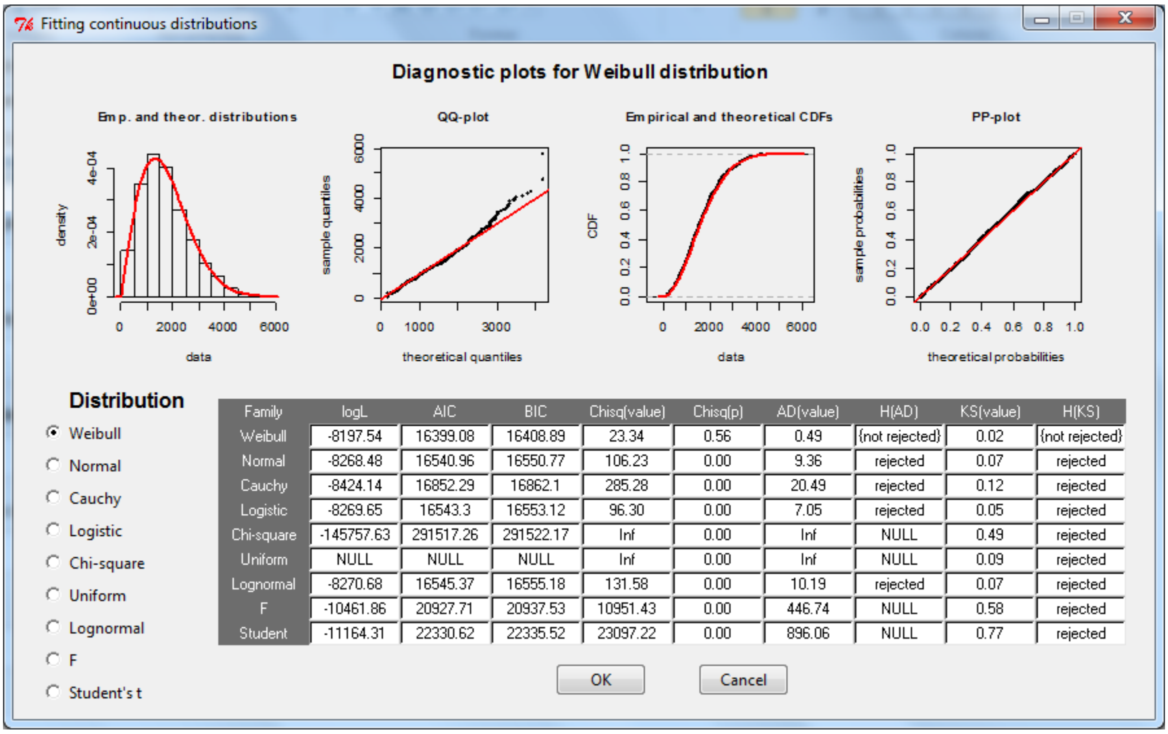


Ahora procedemos a introducir los datos que debe tener como inputs:

Una vez tengamos todos los datos establecidos procedemos a clicar en el botón situado en el margen superior izquierdo "Ajustar distribución demanda".



Posteriormente nos aparecerá el siguiente icono en la barra de inicio.

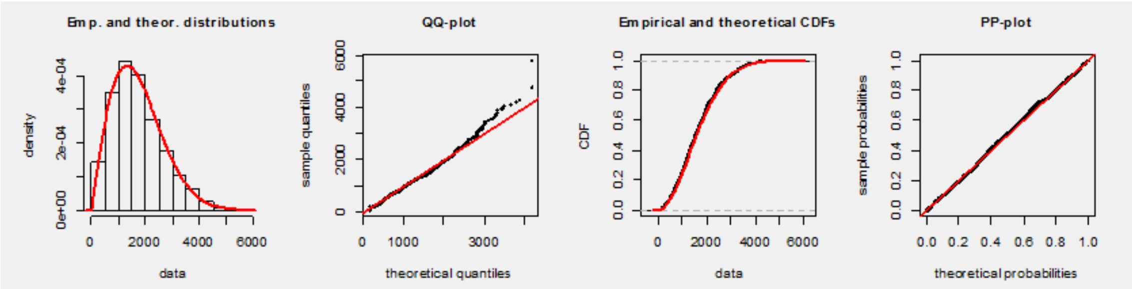


Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Weibull	-8197.54	16399.08	16408.89	23.34	0.56	0.49	{not rejected}	0.02	{not rejected}
Normal	-8268.48	16540.96	16550.77	106.23	0.00	9.36	rejected	0.07	rejected
Cauchy	-8424.14	16852.29	16862.1	285.28	0.00	20.49	rejected	0.12	rejected
Logistic	-8269.65	16543.3	16553.12	96.30	0.00	7.05	rejected	0.05	rejected
Chi-square	-145757.63	291517.26	291522.17	Inf	0.00	Inf	NULL	0.49	rejected
Uniform	NULL	NULL	NULL	Inf	0.00	Inf	NULL	0.09	rejected
Lognormal	-8270.68	16545.37	16555.18	131.58	0.00	10.19	rejected	0.07	rejected
F	-10461.86	20927.71	20937.53	10951.43	0.00	446.74	NULL	0.58	rejected
Student	-11164.31	22330.62	22335.52	23097.22	0.00	896.06	NULL	0.77	rejected

Esta parte del programa nos da una serie de posibles distribuciones como vemos en el margen inferior izquierdo, el mismo programa establece la primera distribución como la que más se aproxime a la muestra seleccionada. Clicando en las distintas distribuciones nos va cambiando los gráficos superiores en los que se ve la diferencia.

En este cuadro se nos muestran varias cosas, comenzaremos comentado los cuatro gráficos que aparecen en la parte superior:

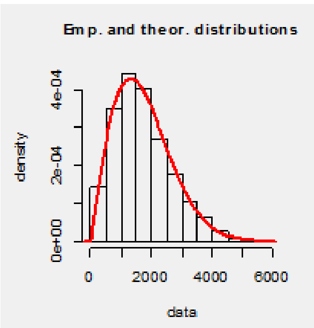
4.6.1.1. Gráficos:

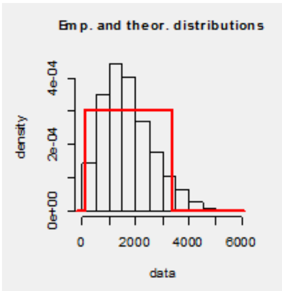


El primer gráfico comenzando por la izquierda refleja un histograma establecido a partir de los datos recogidos, luego en color rojo establece la función que más se aproxima a este tipo de distribución.

Como podemos ver en las siguientes imágenes si cambiamos el tipo de distribución el histograma no varía pero la función sí.

Distribución Weibull. Distribución Normal.



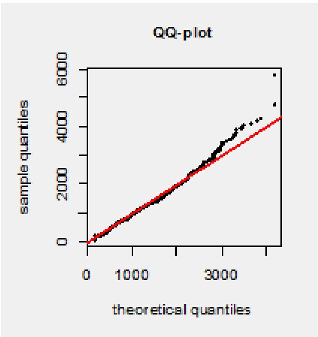


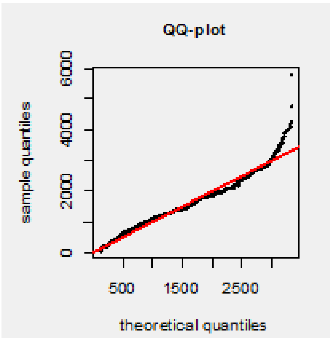
El siguiente gráfico esta relacionado con la regresión lineal, por un lado vemos la distribución de los cuantiles de la muestra seleccionada, y por el otro la recta de regresión lineal que describe para cada tipo de distribución, es grafico es bastante importante ya que podemos ver qué tipo de gráfico nos aproxima mejor.

En estadística, un gráfico Q-Q ("Q" viene de cuantil) es un método gráfico para el diagnóstico de diferencias entre la distribución de probabilidad de una población de la que se ha extraído una muestra aleatoria y una distribución usada para la comparación. Una forma básica de gráfico surge cuando la distribución para la comparación es una distribución teórica.1 No obstante, puede usarse la misma idea para comparar las distribuciones inferidas directamente de dos conjuntos de observaciones, donde los tamaños de las muestras sean distintos.2

Un ejemplo del tipo de diferencias que pueden comprobarse es la no-normalidad de la distribución de una variable en una población. Para una muestra de tamaño n, se dibujan n puntos con los (n+1)-cuantiles de la distribución de comparación, por ejemplo, la distribución normal, en el eje horizontal el estadístico de k-ésimo orden, (para k = 1, ..., n) de la muestra en el eje vertical. Si la distribución de la variable es la misma que la distribución de comparación se obtendrá, aproximadamente, una línea recta, especialmente cerca de su centro. En el caso de que se den desviaciones sustanciales de la linealidad, los estadísticos rechazan la hipótesis nula de similitud.

Distribución Weibull. Distribución uniforme.



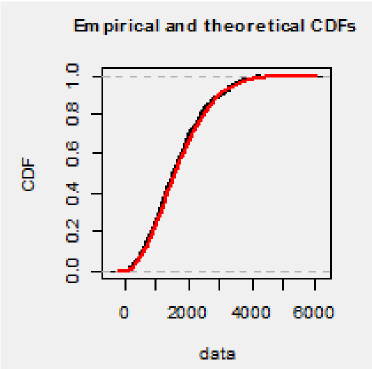


Como vemos en la comparación entre los dos gráfico anteriores tiene menor error el primero que el segundo.

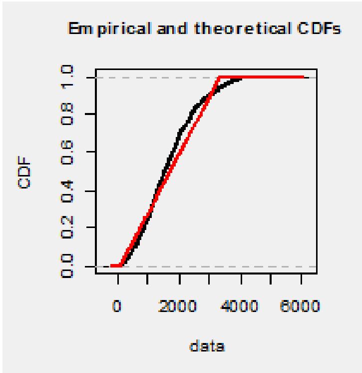
El tercer gráfico se grafica la frecuencia acumulada o frecuencia acumulativa. Que es la frecuencia de ocurrencia de valores de un fenómeno menores que un valor de referencia. El fenómeno puede ser una variable aleatoria que varía en el tiempo o en el espacio. La frecuencia acumulada se llama también frecuencia de no−excedencia. El análisis de la frecuencia acumulada se hace con el propósito de obtener una idea de cuantas veces ocurriría un cierto fenómeno lo que puede ser instrumental en describir o explicar una situación en la cual el fenómeno juega un papel importante.

En las siguientes imágenes mostraremos la diferencia entre dos tipos de distribución y la misma muestra.

Distribución Weibull

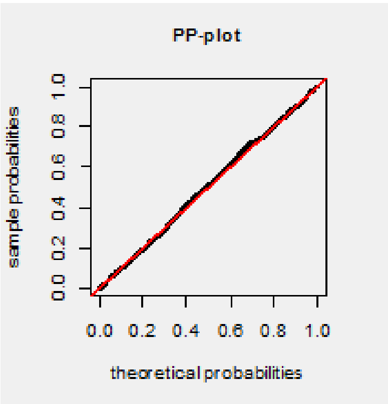


Distribución Normal.

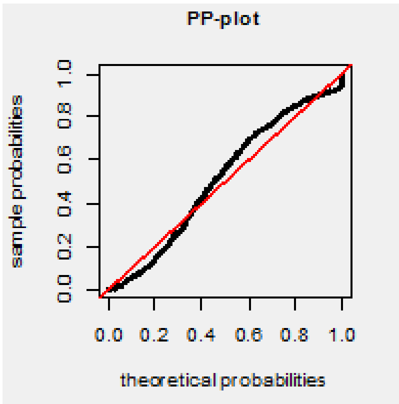


El cuarto gráfico no muestra que en estadística, una parcela P-P (gráfico de probabilidad-probabilidad o ciento por ciento trama) es un gráfico de probabilidad para evaluar cómo de cerca de dos conjuntos de datos están de acuerdo, que traza las dos funciones de distribución acumulada de unos contra otros. Parcelas PP se usan ampliamente para evaluar la asimetría de una distribución.

Distribución Weibull



Distribución Normal.



4.6.1.2. Tabla:

Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Weibull	-8197.54	16399.08	16408.89	23.34	0.56	0.49	{not rejected}	0.02	{not rejected}
Normal	-8268.48	16540.96	16550.77	106.23	0.00	9.36	rejected	0.07	rejected
Cauchy	-8424.14	16852.29	16862.1	285.28	0.00	20.49	rejected	0.12	rejected
Logistic	-8269.65	16543.3	16553.12	96.30	0.00	7.05	rejected	0.05	rejected
Chi-square	-145757.63	291517.26	291522.17	Inf	0.00	Inf	NULL	0.49	rejected
Uniform	NULL	NULL	NULL	Inf	0.00	Inf	NULL	0.09	rejected
Lognormal	-8270.68	16545.37	16555.18	131.58	0.00	10.19	rejected	0.07	rejected
F	-10461.86	20927.71	20937.53	10951.43	0.00	446.74	NULL	0.58	rejected
Student	-11164.31	22330.62	22335.52	23097.22	0.00	896.06	NULL	0.77	rejected

Los parámetros que nos muestra la tabla corresponden a los cálculos mencionados en el Hito I, para estudiar qué tipo de demanda debemos seleccionar debemos considerar los siguientes parámetros.

4.6.1.3. Selección del tipo de la demanda:

Como hemos visto cuando pulsamos el botón Ajuste de la demanda, nos aparecen los gráficos anteriores y su tabla correspondiente, mostrado esto como un catálogo de las diferentes demandas posibles, puede hacer en base a criterios de minimización de error entre los valores reales y los de la distribución de ajuste, o como alternativa en base a un test de bondad de ajuste, en concreto el test de Kolmogorov-Simnorv, los cuales fueron explicados el Hito I, y se encuentran en las siguientes columnas de la tabla. Cabe destacar que el programa propone la mejor distribución en cabeza de la tabla.

Los parámetros que tenemos que tener en cuanta son los marcados en la siguiente tabla.

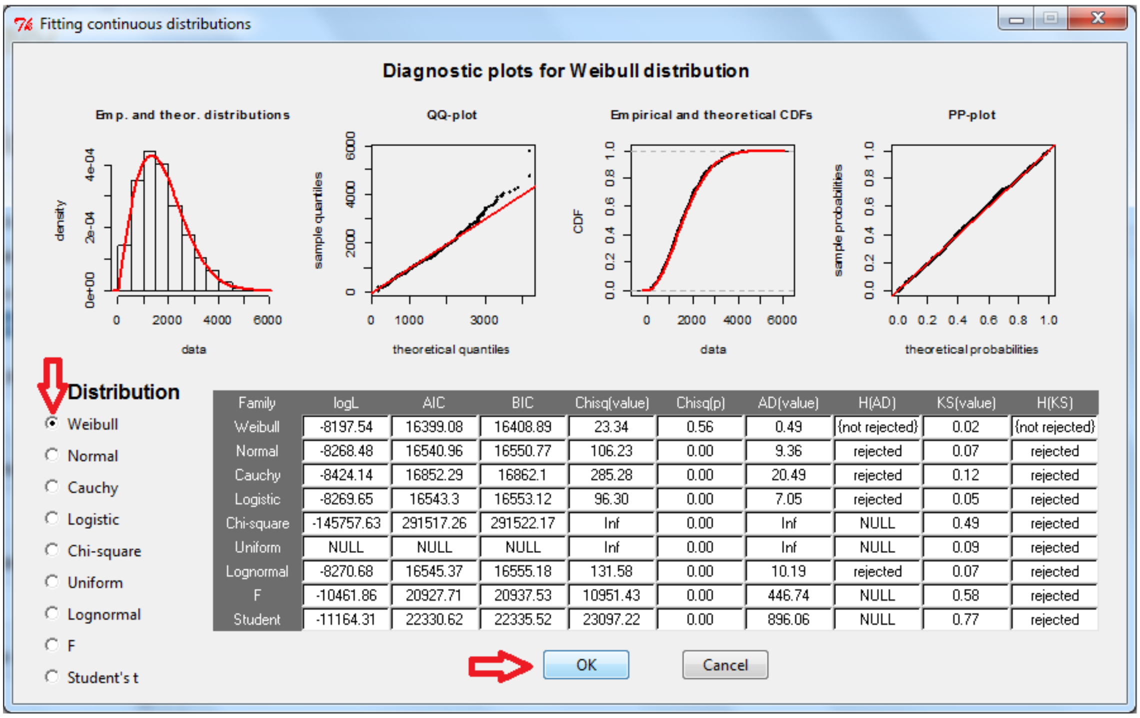
Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Weibull	-8197.54	16399.08	16408.89	23.34	0.56	0.49	{not rejected}	0.02	{not rejected}
Normal	-8268.48	16540.96	16550.77	106.23	0.00	9.36	rejected	0.07	rejected
Cauchy	-8424.14	16852.29	16862.1	285.28	0.00	20.49	rejected	0.12	rejected
Logistic	-8269.65	16543.3	16553.12	96.30	0.00	7.05	rejected	0.05	rejected
Chi-square	-145757.63	291517.26	291522.17	Inf	0.00	Inf	NULL	0.49	rejected
Uniform	NULL	NULL	NULL	Inf	0.00	Inf	NULL	0.09	rejected
Lognormal	-8270.68	16545.37	16555.18	131.58	0.00	10.19	rejected	0.07	rejected
F	-10461.86	20927.71	20937.53	10951.43	0.00	446.74	NULL	0.58	rejected
Student	-11164.31	22330.62	22335.52	23097.22	0.00	896.06	NULL	0.77	rejected

Los dos primeros son el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC), estos dos parámetros están construidos a partir del error acumulado, por lo tanto deberemos tener en cuenta que si queremos comparar gráficas los tamaños de las muestras sean iguales entre ellos.

Las dos últimas columnas están referidas a la prueba de Kolmogorov-Smirnov, en la primera se hace referencia al valor y en la segunda si es una distribución que se acogen dentro de dicha prueba.

Esta parte del funcionamiento es la más comprometida ya que si no tuviéramos unos conocimientos previos de estadística y matemática se nos complicaría dicha deducción, como ya comentamos anteriormente en el Hito I, la selección de un tipo de distribución que refleje la distribución perfectamente de una muestra aleatoria de una demanda queda recogida como la función que más se aproxime a esta, por ello lo que buscamos es el tipo de distribución que más se nos ajuste.

Una vez tenemos claro el tipo de distribución que sigue dicha muestra la seleccíónanos en el lateral de la imagen posterior y pulsamos 'OK' como vemos a continuación.



Ahora en la segunda pestaña 'Ajuste distribución demanda', se nos crea un informe el cual recoge los tipos de distribuciones que más se aproximan y posteriormente en la parte inferior de este informe nos recoge los datos de la tabla anteriormente nombrada.

Los datos más relevantes se encuentran la zona más inferior de la imagen:

```
-------------------------------------------------------------------
Chosen continuous distribution is:  Weibull (weibull)
Fitted parameters are:
      shape       scale
   1.908152 1943.513486
--------------------------------------------------------------------
```

Los cuales reflejan el tipo de distribución seleccionada y en cada caso nos dará los parámetros en los que se basa, en nuestro caso nos da la forma (shape) '1.908152' y la escala (scale) '1943.513486'. Posteriormente mostramos el informe completo como quedará reflejado en pantalla.

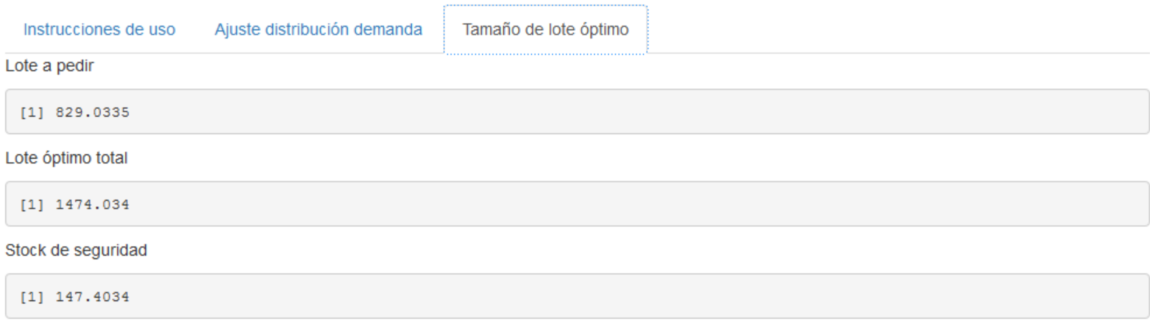
Instrucciones de uso | Ajuste distribución demanda | Tamaño de lote óptimo

```
-------------------------------------------------------------------
Begin fitting distributions...
Normal distribution has been fitted successfully!
Cauchy distribution has been fitted successfully!
Logistic distribution has been fitted successfully!
Warning: Beta distribution couldn't be fitted!
Warning: Exponential distribution couldn't be fitted!
Chi-square distribution has been fitted successfully!
Uniform distribution has been fitted successfully!
Warning: gamma distribution couldn't be fitted!
Lognormal distribution has been fitted successfully!
Weibull distribution has been fitted successfully!
F distribution has been fitted successfully!
Student distribution has been fitted successfully!
Warning: Gompertz distribution couldn't be fitted!
Warning: Triangular distribution couldn't be fitted!
End fitting distributions...
-------------------------------------------------------------------
                 logL       AIC       BIC Chisq(value) Chisq(p) AD(value)        H(AD)
Normal       -8268.48  16540.96  16550.77       106.23     0.00      9.36     rejected
Cauchy       -8424.14  16852.29   16862.1       285.28     0.00     20.49     rejected
Logistic     -8269.65   16543.3  16553.12        96.30     0.00      7.05     rejected
Chi-square -145757.63 291517.26 291522.17          Inf     0.00       Inf         NULL
Uniform          NULL      NULL      NULL          Inf     0.00       Inf         NULL
Lognormal    -8270.68  16545.37  16555.18       131.58     0.00     10.19     rejected
Weibull      -8197.54  16399.08  16408.89        23.34     0.56      0.49 not rejected
F           -10461.86  20927.71  20937.53     10951.43     0.00    446.74         NULL
Student     -11164.31  22330.62  22335.52     23097.22     0.00    896.06         NULL
           KS(value)         H(KS)
Normal          0.07     rejected
Cauchy          0.12     rejected
Logistic        0.05     rejected
Chi-square      0.49     rejected
Uniform         0.09     rejected
Lognormal       0.07     rejected
Weibull         0.02 not rejected
F               0.58     rejected
Student         0.77     rejected
-------------------------------------------------------------------
Chosen continuous distribution is:  Weibull (weibull)
Fitted parameters are:
      shape       scale
   1.908152 1943.513486
--------------------------------------------------------------------
```

En última instancia nos dirigiremos a la tercera pestaña 'Tamaño de lote óptimo', en esta pestaña nos encontraremos con tres salidas numéricas y un gráfico.

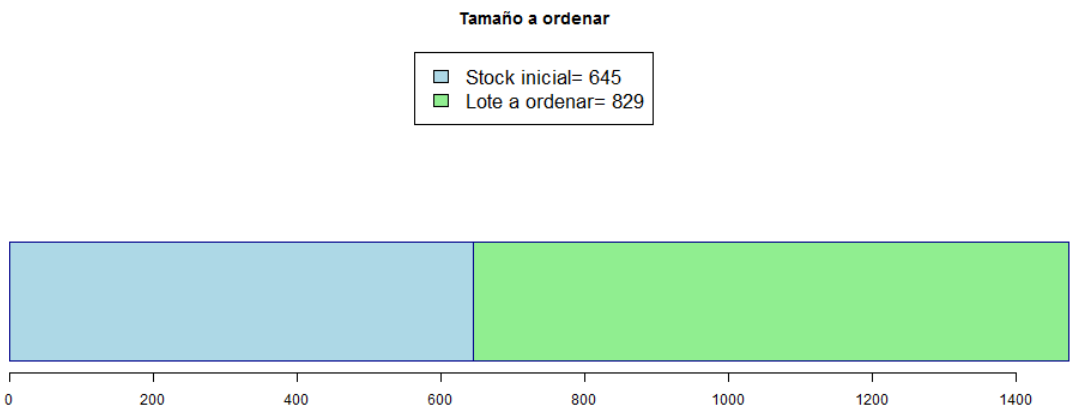
Las salidas numéricas principales están destinadas al valor decimal del lote a pedir lote total y stock de seguridad.

Como vemos en la siguiente imagen:



Aquí se establecen los valores numéricos sin aproximar a la unidad.

Y en la siguiente imaginen nos encontramos con el gráfico.



4.6.2. *Instrucciones de uso:*

Como ya hemos explicado anteriormente el proceso del funcionamiento del programa aquí adjuntamos a modo de resumen la instrucción de uso colocada en la primera pestaña del programa.

Instrucciones de uso | Ajuste distribución demanda | Tamaño de lote óptimo

En primer lugar hay que cargar un fichero xls con la demanda del producto.

Seguidamente introduzca los valores de los costes, nivel de stock inicial y stock de seguridad.

Posteriormente haga click sobre el botón 'Ajustar distribución demanda' y a continuación diríjase a la siguente pestaña.

Espere hasta que le aparezca en pantalla el gráfico con las diferentes opciones de ajuste.

Ahora deberá seleccionar el tipo de distribución que más se ajuste a su distribución, para ello fíjese en los siguienntes parámetros:

El valor (AIC) deberá ser lo más cercano a 0.

El valor (BIC) deberá ser lo más cercano a 0.

El valor del Kolmogórov-Smirnov KS(value) deberá ser lo más cercano a 0.

La prueba de contraste de Kolmogórov-Smirnov H(KS) deberá ser not rejected.

De forma predeterminada el mismo programa nos situará en la parte alta de la tabla la distribución que más se ajuste.

Ahora en la segunda pestaña nos saldrá un informe sobre el ajuste de la demanda.

En la tercera pestaña encontraremos los valores del tamalo del lote optimo total, lote a pedir y stock de seguridad todo ello acompañado de un gráfico.

5. CONCLUSIONES

Una primera conclusión es que hay que fijarse que para el cálculo de las ecuaciones anteriormente mostradas se precisa un conocimiento matemático alto. La base de este trabajo se centra en la optimización de costes y para ello elegimos las herramientas que mejor resultado nos pueden proporcionar. Uno de los aspectos más importantes es entender que debido a que se trata de métodos predictivos la solución mostrada es siempre una aproximación al área del tipo de distribución que indudablemente tenemos que saber establecer.

Una segunda conclusión por mi parte personal tras ver el funcionamiento de OPTINV sentir un una sensación de confort ya que al manejar tantas características de tan complejo significado y ver que funcionan se agradece.

Desde mi punto de vista ha sido un trabajo con bastantes retos de dificultad variable pero que crean una máquina que facilita el trabajo tanto logístico como productivo de las empresas. Personalmente me ha sido de agrado ir avanzando y descubriendo métodos para el cálculo de errores y ver como se correlacionan los datos para que con costes y tipo de distribución podamos crear un método predictivo eficaz.

6. BIBLIOGRAFÍA

http://rstudio-pubs-static.s3.amazonaws.com/24339_4e614e52519d457cb42371c718369123.html

http://www.um.es/ae/FEIR/10/

http://analisisydecision.es/determinar-la-distribucion-de-un-vector-de-datos-con-r/

http://www2.uiah.fi/projects/metodi/280.htm

https://cran.r-project.org/doc/contrib/Chicana-Introduccion_al_uso_de_R.pdf

https://es.wikipedia.org/wiki/Estoc%C3%A1stico

https://es.wikipedia.org/wiki/Funci%C3%B3n_error

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_normal

https://es.wikipedia.org/wiki/Estoc%C3%A1stico

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_log-normal

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_gamma

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_log%C3%ADstica

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_beta

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_exponencial

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_F

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_t_de_Student

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_de_Cauchy

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_%CF%87%C2%B2

https://es.wikipedia.org/wiki/Regresi%C3%B3n_lineal

https://es.wikipedia.org/wiki/Criterio_de_informaci%C3%B3n_bayesiano

https://es.wikipedia.org/wiki/Prueba_de_Kolmog%C3%B3rov-Smirnov

Probabilidad y Estadística (Aplicaciones y métodos) George C.Canavos

Bibliografía

Daniel Peña Sánchez de Rivera Estadística Modelos y Métodos 1.Fundamentos

Relación de documentos

(X) Memoria .... 68 páginas

(_) Anexos .... 3 páginas

La Almunia, a 28 de Junio de 2016

Firmado: [MARIANO CIUTAD BUETAS]