

## METODOS NUMERICOS PARA LA SIMULACION DE UN SISTEMA EXPERTO EN EL ANALISIS ESTADISTICO

### NUMERICAL METHODS FOR THE SIMULATION OF AN EXPERT SYSTEM IN THE STATISTICAL ANALYSIS

**Narcisa J. Salazar, José L. Tinajero y Edgar G. Salazar**

*Grupo de Matemática Computacional, Facultad de Informática y Electrónica, Escuela Superior  
Politécnica de Chimborazo, Panamericana Sur Km. 1, Riobamba Ecuador, nsalazar@epoch.edu.ec,  
jtinajero@epoch.edu.ec, egsalazar@hotmail.com, <http://www.epoch.edu.ec>*

**Palabras clave:** Inferencia Estadística, Sistema Experto, Métodos Numéricos.

**Resumen.** Como una aplicación real y práctica de los métodos numéricos en la Escuela Superior Politécnica de Chimborazo (Riobamba - Ecuador) se pensó en resolver el siguiente problema: Los estudiantes de pre y postgrado tienen dificultades al desarrollar su trabajo final de grado, al momento de usar las tablas de valores de las distribuciones de muestreo para hacer Inferencia Estadística. Para encontrar los intervalos de confianza están limitados a usar los valores que se encuentran en libros, lo cual es un inconveniente ya que al desarrollar temas que se orientan a automatizar y medir las características de calidad de procesos reales, no encontraban los valores que necesitaban para comparar y tomar una decisión. Entonces se pensó que con la ayuda de los métodos numéricos, se podría construir un Programa Informático (Sistema Experto), que simule un análisis para comprobación de hipótesis, para lo cual fue necesario resolver las Ecuaciones Integrales de los estadísticos  $Z$  y  $T$ , creando un método numérico que combine Bisección y Simpson en un solo proceso, tratando de reducir al máximo el costo computacional y el error de aproximación. Este sistema fue aplicado y validado por estudiantes, maestrantes y docentes, llegando a concluir que los tiempos se reducen en un 80%. Con esta aplicación se logró que los estudiantes se motiven y entiendan que los Métodos Numéricos pueden ser muy útiles en su vida estudiantil y profesional.

**Keywords:** Statistical Inference, Expert System, Numerical Methods.

**Abstract.** As a real and practical application of the numerical methods in the Polytechnical School of Chimborazo (Riobamba - Ecuador) is in the following problem: Pre and postgraduate students have the ability to develop their final degree work, when using The tables of values of the sampling distributions to make. In order to find the confidence intervals, they are limited to using the values found in the books, which is an inconvenience and the topics that are oriented to automate and measure the quality characteristics of the real processes, the values have not been determined. They needed to Compare and make a decision. So, I know that with the help of numerical methods, that a Computer Program (Expert System) could be built, simulating an analysis for hypothesis testing, for which it is necessary to solve the Integral Equations of the  $Z$  and  $T$  statistics, creating a numerical method that combines Bisección and Simpson in a single process, trying to reduce to the maximum the computational cost and the approximation error. This system was applied and validated by students and teachers, reaching the conclusion that times are reduced by 80%. This is an application that allows students to be motivated and that Numerical Methods can be very useful in their student and professional life.

## 1 INTRODUCCIÓN

La estadística se ha convertido en una herramienta fundamental para la investigación científica y empírica en todos los campos de la ciencia, sean estos sociales o técnicos. La estadística enfoca en forma directa el gran problema de la toma de decisiones inteligentes y acertadas (Cabrera y Jaramillo, 1995).

Según un estudio realizado en la Facultad de Informática y Electrónica de la Escuela Superior Politécnica del Chimborazo (ESPOCH) en Riobamba Ecuador, se dedujo que los estudiantes de pre y postgrado tienen dificultades al desarrollar su trabajo final de grado, al momento de usar las tablas de valores de las distribuciones de muestreo para hacer Inferencia Estadística. Para encontrar los intervalos de confianza están limitados a usar los valores que se encuentran en libros, lo cual es un inconveniente ya que al desarrollar temas que se orientan a automatizar y medir las características de calidad de procesos reales, no encontraban los valores que necesitaban para comparar y tomar una decisión. Entonces se pensó que con la ayuda de los métodos numéricos, se podría construir un Programa Informático con estructura de Sistema Experto, que simule un Análisis Estadístico para comprobación de hipótesis, con el cual los estudiantes (maestros) podrían tomar rápidamente una decisión eficiente sin mayor dificultad y sin la asesoría directa del profesor o tutor, apoyados solamente en el sistema. Además que se podría enfrentar otro problema que se presenta en los estudiantes de la cátedra de Métodos Numéricos, es el de las aplicaciones reales de estos métodos, lo cual redundaría significativamente en el proceso de enseñanza de esta disciplina. Es necesario mencionar que previo a la ejecución de este proyecto se buscó herramientas informáticas que nos ayuden al proceso de Validación de Hipótesis y no se encontró una herramienta que se acople totalmente a nuestra realidad.

Cabe recalcar que para el desarrollo de este sistema se contó con la colaboración de docentes y estudiantes de las asignaturas de métodos numéricos, estadística, programación e inteligencia artificial.

## 2 CONTENIDOS TEORICOS

### 2.1 Métodos numéricos

Los métodos numéricos son técnicas matemáticas que nos permiten resolver de forma aproximada problemas matemáticos de mediana o alta complejidad, haciendo una estimación del error cometido (Chapra y Canale, 2015; Nieves y Domínguez, 2014).

En la actualidad los Métodos Numéricos se han convertido en una herramienta fundamental que apoyada en la informática permite solucionar problemas reales en todos los campos de la ciencia, sean estos sociales o técnicos (Martínez, 2006). El aprendizaje de los Métodos Numéricos es muy importante para que los estudiantes exploren algoritmos que permitan encontrar soluciones en forma aproximada con un mínimo error, y observen el campo amplio de aplicaciones en diferentes áreas del conocimiento (Bevilacqua y Bini, 1995; Singh y Kushwaha, 2016).

Para resolver el problema de Inferencia Estadística es necesario calcular raíces de ecuaciones integrales de una elevada complejidad, razón por la cual afrontaremos el problema utilizando técnicas de integración numérica y técnicas para hallar raíces de ecuaciones.

### Raíces de ecuaciones

Para hallar las raíces de ecuaciones se cuenta con varios métodos numéricos, entre los

que se pueden citar el método de bisección, el método de Newton, el método de la falsa posición, etc., cada uno de ellos tienen ventajas y desventajas, algunos son más rápidos como el de Newton pero su aplicación puede presentar problemas ya que usa la derivada de la función, lo que implica imponer condiciones de derivabilidad. En el proyecto no se utilizó este método por la complejidad de las ecuaciones (Ec. 3 y Ec. 4), por lo que se optó por el método de bisección, que de las pruebas que se realizó se dedujo que encontraba las soluciones en un tiempo aceptable y con un error despreciable. Este método se detalla a continuación.

Hallar una solución de la ecuación  $f(x)=0$  en el intervalo  $[a,b]$ , con un error máximo de  $e$ . Para aplicar el método de bisección es necesario que  $f(a)$  y  $f(b)$  tengan signos opuestos, este es un método de búsqueda incremental en el que el intervalo se divide siempre en dos. Si la función cambia de signo sobre un intervalo, se evalúa el valor de la función en el punto medio. La posición de la raíz se determina situándola en el punto medio del subintervalo dentro del cual ocurre un cambio de signo. El proceso se repite hasta obtener una mejor aproximación (Chapra y Canale, 2015; Nieves y Domínguez, 2014).

#### **Algoritmo de Bisección**

```
Repeat
   $c=(a+b)/2$ ;
   $signo=f(a)*f(c)$ ;
  if  $signo < 0$  then  $b=c$ ;
  else  $a=c$ ;
   $error=abs(b-a)/2$ 
Until  $error < e$ .
```

#### **Integración numérica (método de Simpson)**

De igual forma para resolver problemas de integración definida se cuenta con varios métodos o formulas numéricas, en este proyecto se utilizó el método de Simpson (1/3), que se obtiene aproximando el integrando mediante un polinomio de interpolación de tres puntos

$$I = \int_a^b f(x) dx \approx \int_a^b P_2(x) dx \quad (1)$$

Lo que nos conduce a la siguiente formula

$$I \cong \frac{(b-a)}{6} [f(a) + 4f(c) + f(b)] \quad (2)$$

Donde  $c$  es el punto medio entre  $a$  y  $b$ . Se escogió este método ya que se realizó un análisis del error producido por esta fórmula y se dedujo que en la mayoría de casos en los que se necesita para el sistema estadístico era despreciable, excepto en el caso en el que el intervalo de integración es mayor a uno, por lo que en esos casos se optó por abrir el dominio de integración y aplicar la aditividad del integral (Singh y Kushwaha, 2016).

#### **Algoritmo de Simpson (Aplicación Múltiple)**

```
 $h=(b-a)/((Trunc(b-a))+1)$ 
Repeat
   $m=a$ ;  $n=a+(h/2)$ ;  $r=a+(h)$ 
   $Integral=Integral+((r-m)/6)*(f(m)+4*f(n)+f(r))$ ;
```

$$a=r;$$

$$\text{Until } a \geq b;$$

## 2.2 Estadística Inferencial

La inferencia estadística es la parte de la estadística matemática que se encarga del estudio de los métodos para la obtención del modelo de probabilidad (forma funcional y parámetros que determinan la función de distribución) que sigue una variable aleatoria de una determinada población, a través de una muestra (parte de la población) obtenida de la misma para sacar conclusiones generales (Montgomery y Runger, 1996).

### Estadístico Z

Sin lugar a dudas, la distribución más utilizada para modelar experimentos aleatorios es la distribución normal o Gaussiana Z. Para realizar inferencia con este estadístico es necesario resolver la siguiente Ecuación Integral

$$f(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt - \alpha = 0, \quad (3)$$

donde  $\alpha$  es el nivel de significancia (margen de error admisible) (Canavos, 1994; Montgomery y Runger, 1996).

### Estadístico T

El estadístico T es una distribución normalizada que se utiliza en los casos en el que la población es pequeña y se desconoce la desviación estándar de la muestra, el parámetro que caracteriza a este estadístico son los grados de libertad (tamaño de la muestra menos uno). Para realizar inferencia con este estadístico es necesario resolver la siguiente Ecuación Integral

$$\frac{\Gamma[(v+1)/2]}{\sqrt{\pi v} \Gamma(v/2)} \int_{-\infty}^{t_{1-\alpha, v}} [1 + (t^2/v)]^{-(v+1)/2} dt = 1 - \alpha, \quad (4)$$

donde

$$\Gamma(n) = \int_0^{\infty} u^{n-1} e^{-u} du, \quad \text{para } n > 0 \quad (\text{Función Gama}) \quad (5)$$

(Canavos, 1994; Montgomery y Runger, 1996)

## 2.3 Sistema Experto

Un Sistema Experto puede ser definido como un sistema basado en conocimiento que emula la competencia de un experto, dentro de un campo de aplicación bien delimitado. También puede ser definido como un conjunto de programas de computadora cuyo comportamiento es similar al que tendría un experto humano en determinado campo (Vílchez, 2007).

La tarea más importante a la hora de desarrollar un sistema experto es la representación del conocimiento, es decir, formalizar y estructurar los objetos y relaciones entre ellos en la base del conocimiento. Además es necesario idear el mecanismo de inferencia (razonamiento) que refleje lo más fielmente posible cómo piensa o actúa un experto. Este mecanismo, que se denomina motor de inferencia, es fundamental en el sistema (Plus, 1999). El método que se utilizó para el Sistema de Análisis Estadístico (SIAE) fue Encadenamiento

hacia adelante, ya que el motor de inferencia parte de los hechos para llegar a los resultados (Russel y Norving, 1998).

*SI premisa ENTONCES consecuencia.*

Para el caso que nos interesa en este proyecto, el sistema daría un apoyo a los estudiantes y profesionales que deseen probar una hipótesis, la ayuda del sistema sería principalmente en los cálculos matemáticos, en los estadísticos a utilizar y en los pasos a seguir para aceptar o rechazar la hipótesis.

### 3 PROPUESTA DE MODELOS NUMERICOS PARA EL SISTEMA

Para la creación de los procesos que resuelven las ecuaciones de los estadísticos planteados es necesario combinar los métodos de Bisección y Simpson en un solo proceso para la distribución  $Z$  y otro para la distribución  $T$ . Además que para calcular la función  $\Gamma$ , se debe realizar una eliminación progresiva de factores.

#### 3.1 Distribución Z

Proceso EstadísticoZ

```
{ Function DistriZ(z:real):real;
  { p:=-4; // Para valores menores a (-4) la función integrada es menor a 0.00001
    h:=abs(z-p)/(trunc(z-p)+1);
    S=0;
    do
      { a:=p; b:=p+h; c:=(a+b)/2;
        S:=S+((b-a)/6)*(exp(-a*a/2)+4*exp(-c*c/2)+exp(-b*b/2));
        p:=b;
      }
    while (p<z);
    S:=(1/sqrt(2*pi))*S;
    DistriZ:=S-alfa;
  }
Procedure Biseccion1;
{ x0:=-4; x1:=4;
  do
    { m:=(x0+x1)/2;
      Signo:=DistriZ(x0)*DistriZ(xm);
      If Signo>=0 then x0:=xm
      Else x1:=xm;
      epsilon:=abs(x0-x1);
    }
    while epsilon>0.001;
  }
}
```

#### 3.2 Distribución T

Proceso EstadísticoT

```
{ Function f(z,v:real):real;
  {
    f:=exp(-(v+1)/2) * ln(1+z*z/v);
  }
};
```

```

Function Gama(w:integer):real;
{ y:=(w+1)/2; z:=w/2;beta:=1;
  y:=y-1; z:=z-1;
  do
    { beta:= beta*(y/z);
      y:=y-1; if z<>1 then z:=z-1;
    }
  while y>=0;
  if y<0 then beta:=beta*sqrt(pi)
  else beta:=beta*(1/sqrt(pi));
  Gama:=beta;
}
Function DistriT(t:real):real;
{ p:=-4; // Para valores menores a (-4) la función integrada es menor a 0.00001
  h:=abs(t-p)/(trunc(t-p)+1);
  S=0;
  do
    { a:=p; b:=p+h; c:=(a+b)/2;
      S:=S+((b-a)/6)*(f(a,v)+4*f(c,v)+f(b,v));
      p:=b;
    }
  while (p<t);
  Gam=Gama(v)*(1/sqrt(pi*v));
  S:=Gam*S; DistriT:=(1-alfa-S);
}
Procedure Biseccion2;
{ x0:=1.28; x1:=5.85; //Dominio de existencia de la solución
  do
    { xm:=(x0+x1)/2;
      Signo:=DistriT(x0)*DistriT(xm);
      If Signo>=0 then x0:=xm
      Else x1:=xm; epsilon:=abs(x0-x1);
    }
  while (epsilon<0.0005)
}
}

```

#### 4 SISTEMA INFORMATICO PARA EL ANALISIS ESTADISTICO (SIAE)

SIAE es un sistema diseñado para brindar un asesoramiento académico de alto nivel, lo que permitirá a los estudiantes notar las aplicaciones de los métodos numéricos y a desarrollar su capacidad científica en el área de la Inferencia Estadística. Para el desarrollo del sistema se utilizó el paradigma de Programación Orientado a Objetos.

Este sistema fue aplicado y validado por estudiantes, maestrantes y docentes de la ESPOCH. Para validar el sistema SIAE se utilizó la prueba Estadística Z obteniendo como resultado que existe una diferencia significativa de un 80% en los tiempos de respuesta entre resolver un problema de forma manual y mediante la aplicación informática.

##### 4.1 Interfaz del Sistema

Las interfaces deben ser amigables y en lo posible similares a las aplicaciones Windows, es decir, los botones, las ventanas, etc. deben hacer funciones parecidas a las que realiza un

sistema clásico de Microsoft, La implementación se realizó con C#.

Al iniciar la aplicación tenemos una pantalla de bienvenida con una imagen fractal, la cual representa la relación que existe entre la estadística y el mundo real (Figura 1).

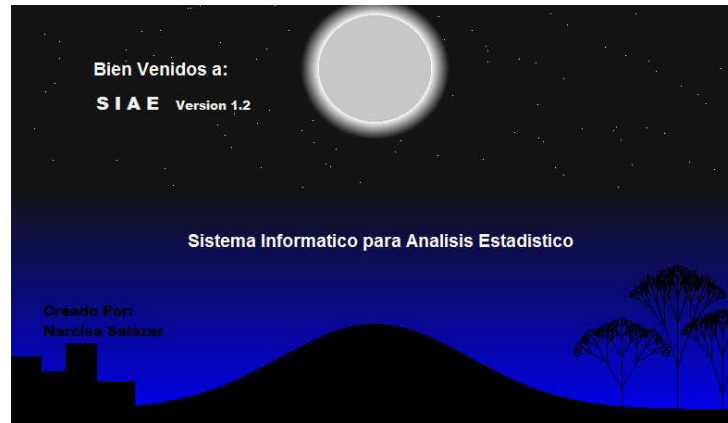


Figura 1: Venta de entrada al sistema

En la pantalla principal (Figura 2) se tienen opciones que inicialmente nos pide escoger si el análisis es de un conjunto de datos o dos. Luego se escoge el tipo de desviación poblacional o muestral. En el siguiente botón pide ingresar los datos que pueden estar tabulados o para tabular, en esta opción se ingresa uno a uno los datos, en este momento actuará el sistema inteligente sugiriendo el estadístico más adecuado. Luego pide ingresar el nivel de significancia variando del 1 al 10 por ciento, una vez ingresado este valor se debe escoger el tipo de análisis si es a una cola derecha, izquierda o a dos colas. En el cuadro de texto se escribe la hipótesis alternativa. Finalmente se tiene el botón que habilita una pantalla de conclusiones finales. Además cuenta con un botón de ayuda de cómo usar el sistema y un botón con ejemplos de Inferencia Estadística.



Figura 2: Venta principal

## 4.2 Resultados del sistema

Los resultados más importantes del sistema se presentan en dos ventanas, la primera en la que indica el intervalo de confianza (Figura 3), y la segunda en la que se presenta la conclusión final respecto a la hipótesis planteada, utilizando un gráfico del estadístico y un resumen que será lo más didáctico y claro para el usuario (Figura 4).

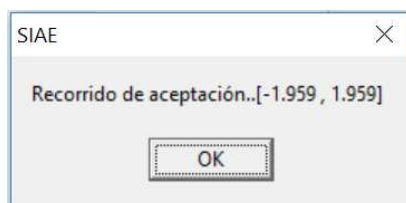


Figura 3: Intervalo de confianza

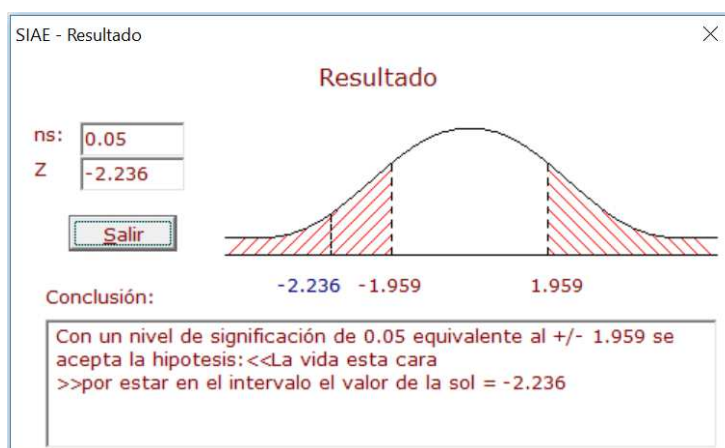


Figura 4: Conclusión final

## 5 CONCLUSIONES

El Sistema informático SIAE fue muy útil para apreciar la importancia de la enseñanza de métodos numéricos en un problema práctico y real que tenía la Facultad de Informática y Electrónica de la ESPOCH, en Riobamba Ecuador, además que fue de gran ayuda en el proceso de Enseñanza – Aprendizaje de la Inferencia Estadística. Se pudo apreciar que los métodos numéricos se pueden combinar entre sí para resolver problemas más generales o complejos. Sirvió como una herramienta de apoyo para los estudiantes de Posgrado de la ESPOCH en un gran número de problemas de investigación que ellos realizaban para poder obtener su título de Magister.

Un resultado interesante de este sistema que no estuvo dentro de su concepción es que debido a su forma de inferir el conocimiento fue utilizado para validar algunos modelos matemáticos, que sirvieron para la creación de simuladores en ciertos departamentos de producción en la ESPOCH y fuera de ella.

## REFERENCIAS

- Bevilacqua, R., and Bini, D., *Introduzione alla matematica computazionale*, ZANICHELLI, 1995.
- Cabrera, C., and Jaramillo, J., *Estadística Inferencial*, UTPL, 1997.



- Canavos, G., *Probabilidad y Estadística, aplicaciones y métodos*, McGraw-Hill 1994.
- Chapra, S., and Canale, R., *Métodos Numéricos para Ingenieros*, McGraw-Hill, 2015.
- Martínez, F., Matemática computacional: Un nuevo pilar para el desarrollo científico y tecnológico, *Revista MAT (Matemáticas en la Frontera)*, 11:102-115, 2006.
- Montgomery, D., and Runger, G., *Probabilidad y Estadística, aplicadas a la ingeniería*, McGraw-Hill, 1996.
- Nieves, A., and Domínguez, F., *Métodos Numéricos aplicados a la ingeniería*, Grupo Editorial Patria, 2014.
- Plus, J., *Sistemas Expertos: Un enfoque pragmático*, UNR, 1999.
- Russel, S., and Norving, P., *Inteligencia Artificial, un enfoque moderno*, Prentice-Hall, 1998.
- Singh, N., and Kushwaha, N., *Computational Methods for Physics and Mathematics*, Alfa Science International, 2016.
- Vílchez, E., Sistemas Expertos para la enseñanza y el aprendizaje de la matemática en la educación superior. *Cuadernos de investigación y formación en educación matemática*, 2007, Año 2, Número 3, pp. 45-67.