

# UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

**Tema:** Herramienta para el análisis estadístico básico de los datos geológicos -  
mineros de los pozos de perforación.

*Trabajo de diploma para optar por el título de  
Ingeniero en Ciencias Informáticas.*

**Autor:** Leandro Miguel Montero Alonso

**Tutores:** Ing. Rosa Elena Peña Lallanilla

Ing. Milenis Fernández Díaz

La Habana, Junio/2013

“Año 55 de la Revolución”



*“No se puede hacer grandes cosas sin grandes amigos.”*

*A toda mi familia por el apoyo, la confianza que han depositado en mí y el amor que me han dado en todo momento.*

*A todos mis amigos por pasar tan agradables y a veces no tan agradables momentos, pero siempre juntos. En especial a Rafa, Lary y Yendry, los tres mosqueteros.*

*A mis tutoras por ser ejemplos de sacrificio y responsabilidad, y guiarme durante la realización de este trabajo.*

*A los profesores y estudiantes del proyecto que siempre tuvieron la gentileza de ayudarme en el momento que lo necesitaba.*

*A todas estas personas especiales para mí, quiero darles las gracias por su ayuda incondicional, su apoyo y dedicación.*

*Con mucho cariño a mis padres, hermano, tíos, abuelos y primos, por ser las personas que me han alentado y apoyado en todo momento. Con ustedes comparto mi triunfo, ya que a ustedes se los debo.*

Declaro por este medio que el compañero Leandro Miguel Montero Alonso con carné de identidad 89032838581, es el autor de este trabajo y que autoriza a la Universidad de las Ciencias Informáticas para hacer uso del mismo en su beneficio, así como los derechos patrimoniales con carácter exclusivo.

Para que así conste, se firma la presente declaración jurada de autoría en La Habana a los \_\_\_\_ días del mes \_\_\_\_ del año \_\_\_\_\_.

\_\_\_\_\_  
Leandro Miguel Montero Alonso

**Autor**

\_\_\_\_\_  
Ing. Milenis Fernández Díaz

**Tutor**

\_\_\_\_\_  
Ing. Rosa Elena Peña Lallanilla

**Tutor**

Tutor: Ing. Rosa Elena Peña Lallanilla

Formación Académica: Ingeniero en Ciencias Informáticas (Julio/2012)

Centro Laboral: Universidad de las Ciencias Informáticas (UCI).

Correo Electrónico: [relallanilla@uci.cu](mailto:relallanilla@uci.cu)

Tutor: Ing. Milenis Fernández Díaz

Formación Académica: Ingeniero en Ciencias Informáticas (Julio/2011)

Centro Laboral: Universidad de las Ciencias Informáticas (UCI).

Correo Electrónico: [mfdiaz@uci.cu](mailto:mfdiaz@uci.cu)

Las empresas y compañías geomineras del país realizan trabajos de estudios, exploración y evaluación geológica – minera, cuyo objetivo es encontrar y evaluar de forma detallada los yacimientos de minerales útiles para satisfacer las demandas de la sociedad. En la exploración, resulta difícil identificar diversas características de las muestras tomadas de los pozos de perforación como por ejemplo: valores atípicos, concentración y distribución en el yacimiento. Además, el desarrollo de un proyecto minero en yacimientos nacionales requiere de la toma de decisiones complejas que guíen las operaciones, decisiones que se fundamentan en el cálculo de estadígrafos, en la representación visual de los resultados obtenidos y en la caracterización estadística de las variables de interés, proceso que se realiza de forma manual y por lo tanto resulta lento e impreciso. Por tal motivo surge la necesidad de desarrollar una herramienta de estadística básica para dotar a la industria minera, de un medio de trabajo automatizado que le permita analizar los datos geológico – mineros. Para el desarrollo de la herramienta se utilizaron diversas tecnologías libres como el entorno de desarrollo integrado QtCreator v2.5.0, el sistema manejador de base de datos PostgreSQL 9.1 y la biblioteca de graficación QCustomPlot 1.0.0, guiado el proceso de desarrollo por la metodología OpenUP. De esta manera, con el despliegue de la herramienta se genera un mejor desempeño en las labores de las empresas y compañías geomineras en cuanto al análisis estadístico de los datos geológico – mineros.

**Palabras clave**

Análisis estadístico, estadística básica, estadígrafos y exploración.

<b>INTRODUCCIÓN.....</b>	<b>1</b>
<b>CAPÍTULO 1: FUNDAMENTOS TEÓRICOS DEL ANÁLISIS ESTADÍSTICO BÁSICO DE LOS DATOS GEOLÓGICO - MINEROS DE LOS POZOS DE PERFORACIÓN.....</b>	<b>6</b>
Análisis estadístico básico de los datos geológico – mineros de los pozos de perforación .....	6
Análisis exploratorio de los datos .....	7
Estadística.....	8
Métodos Estadísticos en la Minería .....	9
Población, muestra y muestreo.....	10
Estadígrafos de medidas de forma, de posición y de dispersión más importantes para el tratamiento de los datos geológico – mineros .....	11
Estadígrafos que no serán implementados .....	15
Gráficos más utilizados para la representación de la distribución de los datos geológico – mineros .....	16
Distribuciones.....	16
Métodos gráficos para visualizar las características del conjunto de datos .....	18
Análisis de soluciones existentes .....	20
Metodología y tecnologías a utilizar en el desarrollo de la herramienta para el análisis exploratorio de los datos geológico – mineros de los pozos de perforación .....	22
Metodología de desarrollo de software .....	22
Lenguaje de Modelado .....	24
Herramienta CASE .....	25
Lenguaje de programación .....	26
Herramienta del Entorno de Desarrollo Integrado .....	26
Sistema Gestor de Base de Datos.....	27
Biblioteca numérica .....	27
Biblioteca de graficación .....	29
Conclusiones parciales .....	30
<b>CAPÍTULO 2: CARACTERÍSTICAS Y DISEÑO DEL SISTEMA.....</b>	<b>31</b>
Modelado del Dominio .....	31
Descripción de las clases del Modelo del Dominio .....	32

Diagrama de clases del Modelo del Dominio .....	33
Descripción del diagrama de clases del Modelo del Dominio .....	34
Requisitos.....	34
Requisitos Funcionales .....	34
Requisitos No Funcionales .....	36
Modelo del Sistema.....	37
Actores del Sistema.....	38
Casos de Uso del Sistema.....	38
Diagrama de casos de uso del sistema.....	39
Descripción textual de los casos de uso del sistema.....	39
Modelo de Diseño .....	42
Arquitectura de la herramienta .....	43
Patrones de diseño.....	44
Diagrama de clases del diseño.....	46
Descripción de las clases del diseño.....	47
Conclusiones parciales .....	50
<b>CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBAS .....</b>	<b>51</b>
Modelo de implementación .....	51
Diagrama de despliegue .....	51
Componente .....	52
Diagrama de componentes .....	52
Modelo de pruebas .....	53
Caso de prueba .....	53
Procedimiento de prueba .....	55
Resultados de las pruebas.....	62
Conclusiones parciales .....	63
<b>CONCLUSIONES GENERALES .....</b>	<b>64</b>
<b>RECOMENDACIONES.....</b>	<b>65</b>
<b>BIBLIOGRAFÍA.....</b>	<b>66</b>

Tabla 1 Estadígrafos .....	12
Tabla 2 Actor del Sistema.....	38
Tabla 3 Descripción del Caso de Uso del Sistema Calcular estadígrafo .....	41
Tabla 4 Descripción de la clase del diseño AnalisisEstadistico. ....	49
Tabla 5 Caso de prueba del CUS Calcular estadígrafos .....	54
Tabla 6 Procedimiento de prueba del CUS Calcular estadígrafos .....	56
Tabla 7 Caso de prueba para el método seleccionDatosAtipicos .....	59
Tabla 8 Caso de prueba para el método seleccionDatosAtipicos .....	59
Tabla 9 Caso de prueba para el método seleccionDatosAtipicos .....	60
Tabla 10 Caso de prueba para el método seleccionDatosAtipicos .....	60
Tabla 11 Caso de prueba para el método seleccionDatosAtipicos .....	61
Tabla 12 Caso de prueba para el método seleccionDatosAtipicos .....	61
Tabla 13 Caso de prueba para el método seleccionDatosAtipicos .....	61
Tabla 14 Resultados de los casos de prueba .....	62

Ilustración 1 Distribución normal .....	17
Ilustración 2 Distribución logarítmica – normal .....	18
Ilustración 3 Diagrama de clases del Modelo del Dominio .....	33
Ilustración 4 Diagrama de Casos de Uso del Sistema.....	39
Ilustración 5 Diagrama de clases del diseño del CUS Calcular estadígrafos .....	47
Ilustración 6 Diagrama de Despliegue .....	52
Ilustración 7 Diagrama de componentes del CUS Calcular estadígrafos .....	52
Ilustración 8 Método seleccionDatosAtipicos .....	57
Ilustración 9 Grafo de flujo de datos para el método seleccionDatosAtipicos .....	58

## Introducción

La revolución científico – técnica en la geología y la minería, que desde hace ya varias décadas compete a todo el mundo, se pone de manifiesto en el incremento vertiginoso de la prospección, exploración, extracción y utilización de la materia prima mineral, dado que es una de las ramas de la industria que más ingresos reporta a países ricos en recursos minerales. Es primordial asegurar la existencia de suficiente mineral en la zona de estudio antes de invertir recursos en el proyecto, puesto que los costos monetarios asociados al desarrollo de estos son extremadamente elevados. Por lo tanto, una de las tareas más importantes en las empresas y compañías geomineras la constituyen los trabajos de estudios, exploración y evaluación geológica - minera de los yacimientos minerales útiles.

La década comprendida entre 1970 y 1980 se caracterizó en el país por una elevada actividad de prospección geológica<sup>1</sup>; dicha industria en 1975 había alcanzado el doble de la producción de 1958; y se planteaban nuevas inversiones en la rehabilitación de las industrias existentes y la construcción de otras. La necesidad de manejar un enorme volumen de datos hizo imprescindible el uso de las técnicas de computación para el control del flujo de datos, los cálculos de reservas, mejorar la información geológica obtenida y la confección del material gráfico. Es por ello que la historia del uso de computadoras en la industria minera cubana tiene sus orígenes en el tratamiento automatizado de la información geológica en los yacimientos de níquel, y a medida que se aumentó el empleo de esta tecnología, se ampliaron sus aplicaciones en otros lugares y hacia otros fines **(Ruiz, 2008)**.

El auge que fue tomando esta rama en la sociedad cubana promulgó la necesidad de un software para el cálculo de las reservas de níquel, necesidad identificada conjuntamente por las plantas Punta Gorda y Las Camariocas, y que da lugar al surgimiento de Níquel. A partir de los requisitos definidos, el software fue desarrollado por la Empresa de Computación del MINBAS EDESAD, conocida hoy como Tecnomática. Con el perfeccionamiento del manejo de las bases de datos y la incorporación del pronóstico y control de la minería se ha ido enriqueciendo este software. Su versión más integral para la prospección geológica del níquel es MICRONIQ. Estas soluciones informáticas son las primeras que se desarrollan con esfuerzo nacional, sin embargo, no se ha logrado integrar en un único sistema computarizado todas las actividades relativas al minado **(Polanco Almanza, y otros, 2001)**.

---

<sup>1</sup> Etapa en la que se buscan minerales aprovechables en una zona determinada.

A principios de los años 90 las empresas geomineras del país comenzaron un esfuerzo para desarrollar un software de minería que extendía lo logrado en el níquel hacia otros tipos de yacimientos, pero la dispersión del equipo nacional de desarrollo y el período especial hicieron que se abortara la tarea. Las primeras versiones de GEMCOM para DOS no superaban el diseño de aquel intento. Al reanimarse la industria del níquel se amplían los esfuerzos para la introducción de la informática en las minas, es cuando surge la disyuntiva de si utilizar un software que siga los estándares mundiales confeccionado en el país o adquirirlo en el mercado. Los especialistas en su totalidad abogaban por desarrollar el software cubano, sin embargo, se inclinó el vector de la decisión hacia la adquisición de los productos informáticos mineros imperantes en el mundo. Estos respondieron a las urgencias, pero impusieron una dependencia tecnológica, el oneroso pago de versiones, licencias, capacitación y servicios, y el compromiso con el derrotero futuro de las compañías productoras de estos sistemas **(Ruiz, 2008)**.

La principal fuente de interés del buen desarrollo de las actividades mineras recae sobre “los recursos minerales sólidos”. En los proyectos mineros la identificación de diversas características de los recursos minerales como por ejemplo: valores atípicos, saltos y discontinuidades, concentración y la forma de distribución en el yacimiento, se realiza a partir de un amplio volumen de datos geológico – mineros, provenientes de las campañas de exploración. A esto se le suma que se debe manejar un número considerable de variables tales como: coordenadas de ubicación, longitud, litología, leyes de cada mineral, entre otras. Estas características determinan en gran medida el valor industrial de un yacimiento mineral.

El desarrollo de los proyectos mineros en los yacimientos nacionales, donde la cantidad de muestras que se toman puede ser superior a 30 000 requiere de la toma de decisiones complejas que guíen las operaciones. Estas decisiones se fundamentan en el cálculo de estadígrafos, en la representación de los resultados a través de gráficos y en la caracterización estadística de las variables de interés a través de reportes que se generen. Este proceso se realiza de forma manual e implica gran cantidad de cálculos sobre una vasta información geológica - minera, lo que puede traer consigo que la estimación de los recursos mineros no se realice correctamente.

La necesidad actual de los recursos minerales y los parámetros de costos e inversiones asociados al desarrollo de un proyecto minero, imponen la aplicación de un conjunto de técnicas, modelos y métodos que permitan analizar todas sus características y finalmente resumir la información y facilitar así su análisis e interpretación, reconociéndose como **problema a resolver** el hecho de: el análisis exploratorio

de los datos geológico – mineros de los pozos de perforación resulta lento e impreciso, debido a la gran cantidad de información geológica – minera a analizar, el número considerable de variables que se manejan y los numerosos cálculos que implica.

El **objeto de estudio** sobre el cual se enfoca la presente investigación, tanto desde el punto de vista teórico como práctico con vista a dar solución al problema planteado, está centrado en: el análisis estadístico básico de los datos geológico - mineros de los pozos de perforación y, como **campo de acción**: el análisis exploratorio de los datos geológico - mineros de los pozos de perforación. El **objetivo general** es implementar una herramienta de estadística básica que permita realizar el proceso de análisis exploratorio de los datos geológico - mineros de los pozos de perforación, procurando **defender la idea** de que: con la implementación de una herramienta de estadística básica, se garantizará mayor rapidez y confiabilidad en el proceso de análisis exploratorio de los datos geológico - mineros de los pozos de perforación.

Para dar cumplimiento al objetivo de la investigación fueron trazadas las siguientes **tareas**:

- Caracterizar el análisis exploratorio de los datos geológico - mineros de los pozos de perforación.
- Caracterizar los estadígrafos de medidas de tendencia central, de medidas de posición y de medidas de dispersión más importantes para el tratamiento de los datos geológico - mineros de los pozos de perforación.
- Describir los gráficos más utilizados para la representación de la distribución de los datos geológico - mineros de los pozos de perforación.
- Seleccionar la metodología y las tecnologías a utilizar en el desarrollo de la herramienta para el análisis exploratorio de los datos geológico - mineros de los pozos de perforación.
- Modelar los procesos de negocio y diseño de la herramienta para el análisis exploratorio de los datos geológico - mineros de los pozos de perforación.
- Implementar la herramienta para el análisis exploratorio de los datos geológico - mineros de los pozos de perforación.
- Validar la herramienta para el análisis exploratorio de los datos geológico - mineros de los pozos de perforación mediante pruebas de software.

Se utilizan los siguientes métodos científicos de la investigación para un mejor entendimiento de la situación existente:

### **Teóricos**

**Análítico - Sintético:** Se realiza un estudio de las teorías presentes en las bibliografías consultadas referentes al análisis estadístico básico de los datos para obtener una síntesis detallada del tema.

**Histórico - Lógico:** Se realiza un estudio de las bibliografías relacionadas con la utilización de los estadígrafos en datos geológico - mineros de los pozos de perforación, obtenidos durante las campañas de exploración, para conocer cuáles de estos son los utilizados en el análisis estadístico.

### **Empíricos**

**Revisión de fuentes bibliográficas:** Se realiza una revisión de las bibliografías relacionadas con el análisis estadístico básico de los datos geológico - mineros de los pozos de perforación, para llegar a entender cómo se realiza este proceso y cada uno de los conceptos asociados al dominio del problema.

### **Resultados**

- Herramienta para el análisis exploratorio de los datos geológico - mineros de los pozos de perforación.
- Documentación correspondiente a la herramienta para el análisis exploratorio de los datos geológico - mineros de los pozos de perforación.

### **Estructura del documento**

El presente documento se encuentra dividido en: introducción, tres capítulos, conclusiones, recomendaciones, bibliografía y anexos. En el primero de los capítulos, se incluyen los conceptos asociados al dominio del problema, la definición del tratamiento estadístico al que se someten los datos geológico - mineros de los pozos de perforación recogidos durante las campañas de exploración, la caracterización de los estadígrafos y gráficos más utilizados para la representación de los datos geológico - mineros, el análisis de las soluciones existentes y por último la selección de la metodología y tecnologías a utilizar en el desarrollo de la herramienta. En el segundo capítulo se realiza una descripción de las características y el diseño de la herramienta a implementar. Por último, en el tercer capítulo se documenta detalladamente los elementos fundamentales referentes a la implementación de la herramienta para

realizar el análisis exploratorio de los datos geológico - mineros de los pozos de perforación, se realizan las pruebas competentes a la herramienta implementada y se presentan los resultados obtenidos del desarrollo del presente trabajo.

## **Capítulo 1: Fundamentos teóricos del análisis estadístico básico de los datos geológico - mineros de los pozos de perforación**

Para desarrollar un proyecto minero resulta necesario conocer la forma del yacimiento, la concentración de mineral en el macizo rocoso, así como el valor de dicho depósito resultante de la evaluación de los beneficios de la explotación. Para definir todas estas características se realizan previamente trabajos de prospección y exploración de los yacimientos minerales. El análisis estadístico de los sondeos permite indicar el comportamiento de las variables que influyen en el proyecto como la distribución de minerales, saltos o discontinuidades. Casi todas las decisiones que se hacen respecto a un proyecto minero, desde la prospección hasta el cierre de la mina, están basadas en los valores obtenidos de la estimación del material muestreado.

En este capítulo se presentan aspectos relacionados con los elementos teóricos fundamentales que sustentan el análisis exploratorio de los datos, se caracteriza el análisis estadístico básico de los datos, los estadígrafos de medidas de tendencia central, de medidas de posición y de medidas de dispersión más importantes para el tratamiento de los datos, se realiza una descripción de los gráficos más utilizados para la representación de la distribución de los datos, se lleva a cabo el análisis de soluciones existentes y por último, la selección de la metodología y las tecnologías a utilizar en el desarrollo de la herramienta.

### **Análisis estadístico básico de los datos geológico – mineros de los pozos de perforación**

Constantemente se recibe datos provenientes de las campañas de exploración minera que resultan de gran utilidad para el proyecto minero en el que se esté trabajando. Estos contribuyen a la toma de decisiones para hacer frente a problemas relacionados con la caracterización de los recursos minerales en los yacimientos cubanos. Es por ello que se necesita un adecuado procesamiento de los datos para arribar a conclusiones certeras que permitan identificar las características de las muestras de los pozos de perforación que son obtenidas del suelo durante las campañas de exploración.

## **Análisis exploratorio de los datos**

El Análisis Exploratorio de los Datos (AED) es un conjunto de técnicas estadísticas que tiene como finalidad examinar los datos, de esta forma el analista obtiene un entendimiento básico de los mismos. El examen previo de los datos es un paso necesario y habitualmente descuidado por parte de los analistas de datos. Las tareas implícitas en dicho examen pueden parecer insignificantes y sin consecuencias a primera vista, pero son una parte esencial de cualquier análisis estadístico **(Salvador Figueras, y otros, 2003)**.

Los M.Sc Edith Alarcón Matutti y César Gutiérrez Villafuerte afirman que el AED consiste en la obtención de medidas de resumen y gráficos para un conjunto de datos, ayuda en la determinación de las técnicas estadísticas a emplear luego, y puede indicar si es necesaria la transformación de alguna variable. Es importante hacer un AED para:

- Evaluación general de los datos.
- Identificación de datos atípicos.
- Descripción.
- Chequeo de supuestos para análisis posteriores.
- Identificar diferencias entre sub - grupos.

Según **(Cruz, 2009)** realizar un análisis estadístico de los datos disponibles o los generados a partir del cálculo de los compósitos<sup>2</sup> (con el objetivo de caracterizar el comportamiento estadístico de las distintas variables en el depósito y en las unidades geológicas que lo integran) es necesario antes de proceder a la estimación de reservas. La organización de los datos cuantitativos y su análisis pueden consumir el 50 % del tiempo necesario para realizar la estimación de reservas, donde el AED está dirigido a resolver las siguientes cuestiones:

- Identificar y eliminar los posibles errores.
- Caracterización estadística de las variables de interés.

---

<sup>2</sup> Procedimiento mediante el cual las muestras de los análisis se combinan en intervalos regulares, que no coinciden con el tamaño inicial de las muestras.

- Identificar y definir los dominios geológicos que requieren un tratamiento independiente durante la estimación de recursos.
- Identificar y caracterizar las muestras con valores extremos (*outliers*).

El cálculo de la estadística básica para las muestras originales y compositadas en cada dominio geológico y el estudio de la distribución de frecuencias de los parámetros constituyen las principales herramientas que posibilitan el AED (**Cruz, 2009**).

### **Estadística**

Una característica fundamental de un proyecto minero es la abundancia de datos. Los especialistas se sienten continuamente bombardeados por estos y deben aprender a leerlos e interpretarlos críticamente. Además de ser capaces de llevar a cabo dichas funciones, deben formular preguntas adecuadas e importantes y generar la información que faciliten las respuestas apropiadas.

Según (**Salazar, 2002**), específicamente la estadística trata de teoremas, herramientas, métodos y técnicas que se pueden usar en la recolección, selección y clasificación de datos; interpretación y análisis de los mismos; deducción y evolución de conclusiones y de su confiabilidad, basada en datos muestrales. Es una ciencia que cubre un extenso campo de aplicación y que se agrupa en dos grandes áreas: la estadística descriptiva y la estadística inferencial, que desempeñan funciones diferentes pero complementarias en el análisis. La estadística descriptiva está basada en técnicas sencillas desde el punto de vista matemático, utilizadas para resumir y describir datos numéricos sin inferir en un grupo mayor a través de gráficos, tablas y medidas de posición, de forma y dispersión. La estadística inferencial es el proceso de hacer predicciones acerca de una población basado en la información de una muestra.

Se considera a la estadística la ciencia de los datos que implica la colección, clasificación, síntesis, organización, análisis e interpretación de la información (**Cassiraga, 2005**). Emplea conocimientos de otras ciencias como la lógica y la matemática, y es una forma razonable de emplear el sentido común y la aritmética para el manejo de los datos en las investigaciones. Explora, describe los datos, condensa y sirve de vehículo para comunicar la información. De manera general, la estadística es un conjunto de técnicas que, partiendo de la observación de fenómenos, permite al investigador obtener conclusiones útiles sobre ellos. La presente investigación se centra en la estadística descriptiva univariada.

## **Métodos Estadísticos en la Minería**

Según el Dr. Jesús Reynaga Obregón el método estadístico consiste en una secuencia de procedimientos para el manejo de datos cualitativos y cuantitativos de la investigación. Dicho manejo tiene como propósito la comprobación, en una parte de la realidad, de una o varias consecuencias verificables deducidas de la hipótesis general de la investigación. Las características que adoptan los procedimientos del método estadístico dependen del diseño de investigación seleccionado para la comprobación de la consecuencia verificable en cuestión.

El método estadístico consta de las siguientes etapas:

- **Recolección:** se obtiene la información cualitativa y cuantitativa señalada en el diseño de la investigación, también conocida como variables.
- **Recuento:** la información obtenida es sometida a revisión, clasificación y cómputo numérico. Consiste en la cuantificación de la frecuencia con que aparecen las diversas medidas en los elementos en estudio.
- **Presentación:** se elaboran los cuadros y los gráficos que permitan efectuar una revisión numérica precisa y facilitar la inspección visual rápida de la información.
- **Síntesis:** la información es resumida en forma de medidas que permiten expresar de forma sintética las principales propiedades numéricas de grandes series o agrupamientos de datos.
- **Análisis:** se realiza la comparación de las medidas de resumen previamente calculadas con el propósito de decidir si existe diferencias significativas entre ellas.

El tratamiento de los datos geológico - mineros con la utilización de la informática ha tenido un notable aumento, teniendo en cuenta que la interpretación de un determinado conjunto de datos puede ser mejorada y detallada, es decir, se pueden elaborar estudios con mayor connotación científica. Con el avance de la informatización en todas las áreas de la actividad humana, los métodos estadísticos que existían desde antaño son más comunes y esenciales como herramienta de interpretación de datos geológico - mineros. Quizás una de las razones más importantes por lo cual sucede esto sea el reconocimiento del alto grado de variabilidad de los atributos estudiados. Otra razón destacada que influye en dicho auge es la necesidad de realizar generalizaciones y predicciones a partir de un número limitado

de datos, en muchos casos, debido a la imposibilidad de acceso directo a los objetivos investigados **(Alfonso Roche, 1989)**.

### **Población, muestra y muestreo**

Gran parte del trabajo de un estadístico profesional se realiza con muestras, debido a que generalmente las poblaciones son demasiado grandes y resulta difícil estudiarlas en su totalidad. Seleccionar una muestra de la población, calcular el estadígrafo a esa muestra, y emplearlo para estimar el parámetro correspondiente de la población es lo más factible, o de lo contrario exigiría demasiados recursos si se estudia la población entera.

Según **(Sironvalle, 2002)** se define a la población, la muestra y el muestreo de la siguiente forma:

*“Población o lote es el conjunto completo de observaciones que deseamos estudiar.”*

*“Muestra es una parte o porción extraída de un conjunto por métodos que permiten considerarla como representativa del mismo.”*

*“Muestreo es la acción de recoger muestras representativas de calidad o condiciones medias de un todo o la técnica empleada en esta selección o la selección de una pequeña parte estadísticamente determinada para inferir el valor de una o varias características del conjunto.”*

La mayoría de las decisiones que se toman en un proyecto minero, desde la prospección hasta el agotamiento de los recursos minerales, o no rentabilidad de la explotación y posterior cierre de la mina, se fundamentan en los valores obtenidos del material muestreado. Una de las operaciones más importantes, tanto en los trabajos de prospección como en los de exploración, es el muestreo, que no es más que un conjunto de trabajos encaminados a determinar la composición cuantitativa y cualitativa de los componentes útiles e impurezas de los minerales. Este conjunto de trabajos (por intermedio de las muestras) permite el estudio de las propiedades físicas y las características técnicas de los minerales **(Sironvalle, 2002)**.

## **Estadígrafos de medidas de forma, de posición y de dispersión más importantes para el tratamiento de los datos geológico – mineros**

Los estadígrafos son las medidas de resumen que se aplican en el análisis estadístico sobre los datos geológico - mineros. Después de haber recogido la información cuantitativa, se ordena y se describe para ser utilizada por estos indicadores, ya que solamente con el ordenamiento y la descripción de los datos, aún resulta insuficiente para el análisis. Estos indicadores permiten hallar un valor numérico que representa a toda la población o muestra en estudio.

Los datos primarios son el conjunto de valores asociados a un atributo de una población muestral. Para el estudio de estos datos es necesario aplicar procedimientos que permitan caracterizar la variación general del atributo, identificar si el mismo presenta preferencias en tomar ciertos valores y caracterizar cuantitativamente el grado de dispersión del atributo alrededor de este valor preferencial (**Alfonso Roche, 1989**).

Después de recoger toda la información correspondiente a la investigación, se presenta una aglomeración de datos y cifras desordenadas, las cuales dicen muy poco sobre el evento natural estudiado. Resulta conveniente distribuir por clases el valor que toman los atributos del evento natural y determinar el número de muestras unitarias<sup>3</sup> que pertenecen a cada clase para un estudio de la población muestral. Una vez realizada la distribución muestral, se dice que los datos primarios están agrupados. Los siguientes métodos numéricos son utilizados para hallar los valores numéricos que representan a toda la población o muestra en estudio:

- **Medidas de posición:** son medidas de un conjunto de datos que proporcionan un valor simple y representativo, que resume un gran volumen de información. Informan acerca de dónde están localizadas las distintas partes de una distribución.
- **Medidas de dispersión:** describen la variabilidad de los valores de los datos. Nos informan sobre cuánto se alejan del centro los valores de la distribución.
- **Medidas de forma:** proveen información acerca de la forma de una distribución (simetría, longitud de las colas, pico).

---

<sup>3</sup> Cada individuo que pertenece a la población muestral es una muestra unitaria

Los estadígrafos que serán implementados se resumen en la siguiente tabla:

Medidas de posición		Medidas de dispersión	Medidas de forma
Localización del centro	Localización de otras partes		
Media aritmética	Mínimo	Rango	Coefficiente de asimetría
Media geométrica	Máximo	Desviación estándar	Coefficiente de Curtosis
Mediana	Cuartiles	Varianza	
Sumatoria		Coefficiente de variación	
Centro de amplitud		Desviación media o absoluta	
		Rango intercuartil	

Tabla 1 Estadígrafos

En teoría del muestreo de minerales se utilizan las nociones de exactitud y precisión, que en términos estadísticos corresponden respectivamente a la **media**, la cual debe ser insesgada (exactitud) y a la **varianza** del error, la cual debe ser pequeña (precisión) (**Sironvalle, 2000**).

La **media aritmética** es el promedio de los  $n$  valores medidos. Posee el inconveniente de que es sensible a la presencia de valores extremos en los datos.

La **media geométrica** es el resultado de la raíz  $n$  - ésima de la multiplicación de un conjunto de  $n$  datos.

La **mediana** de un conjunto de datos es el punto central de los valores observados si se organizan en orden ascendente y la cantidad es impar, en caso de ser par la cantidad de muestras, es la media

aritmética de los dos puntos centrales. No se basa en la magnitud de los datos, como la media aritmética, sino en la posición central que ocupa en el orden de su magnitud, dividiendo la información en dos partes iguales, dejando igual número de datos por encima y por debajo de ella (**Salazar, 2002**).

Si la distribución se ajusta al modelo normal (**ver Distribuciones**), la media aritmética y la mediana coinciden, en caso de que la distribución se ajuste al modelo logarítmico – normal (**ver Distribuciones**), la mediana es mayor que la que la media aritmética (**Cassiraga, 2005**).

La mediana es un buen estadígrafo cuando la distribución es muy asimétrica porque no se ve tan afectada como la media por valores extremos.

Por **sumatoria** se entiende a la cantidad de datos existente en el conjunto.

El **centro de amplitud** es el valor medio que queda entre los valores mínimo y máximo. Está dado por la siguiente fórmula:

$$C_a = \frac{X_{m\acute{a}x} + X_{m\acute{i}n}}{2}$$

El **mínimo** de una variable es el valor para un porcentaje del 0% y el **máximo** es el valor para un porcentaje del 100%.

El **rango** de un conjunto de datos es la diferencia entre el número mayor y el número menor del conjunto.

Los **cuartiles** se definen como los tres valores que dividen la distribución en cuatro partes iguales. El primer cuartil deja el 25% de la información por debajo de él, y el 75% por encima, el segundo cuartil, al igual que la mediana, divide la información en dos partes iguales, y por último el tercer cuartil deja el 75% por debajo de sí, y el 25% por encima.

El rango es muy influenciado por valores extremos, esto en estadística se resuelve con el rango intercuartil, el cual analiza la situación del medio de la distribución. Por lo tanto, el **rango intercuartil** se define como la diferencia entre el tercer cuartil y el primer cuartil.

La **desviación estándar** se calcula como la raíz cuadrada de la varianza. Esta medida con frecuencia se prefiere en lugar de la varianza debido a que sus unidades son las mismas que la variable que se estudia. Esta dada por la siguiente fórmula:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$$

La **varianza** es la desviación cuadrática promedio de los datos con respecto a la media aritmética. Esta medida es sensitiva a valores extremos. Esta dada por la siguiente fórmula:

$$\vartheta = \sigma^2$$

La **desviación media o absoluta** mide la distancia absoluta promedio entre cada uno de los datos y el parámetro que caracteriza la información. Usualmente se considera la desviación media con respecto a la media aritmética. Este indicador muestra que tan disperso se encuentran un conjunto de datos con respecto a un punto de concentración. La desviación media es una medida de dispersión bastante objetiva, es decir, mientras más dispersos están los datos, mayor es la desviación media. Esta dada por la siguiente fórmula:

$$D_m = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n}$$

El **coeficiente de asimetría** valora la simetría de una distribución. Debido a la manera en que se calcula este coeficiente los valores del mismo pueden estar afectados por la presencia de valores extremos en los datos (**Cruz, 2009**). Está dado por la siguiente fórmula:

$$CA = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\sigma^3}$$

- Si CA = 0 presenta una distribución simétrica.
- Si CA > 0 la cola de la distribución es hacia la derecha y se dice que ésta es asimétrica positiva.
- Si CA < 0 la cola de la distribución es hacia la izquierda y se dice que ésta es asimétrica negativa.

El **coeficiente de variación** se usa como una alternativa al coeficiente de asimetría, principalmente para distribuciones en las cuales todos sus valores son positivos y cuya asimetría es también positiva, aunque puede ser utilizado para distribuciones con asimetría negativa. Este coeficiente es una medida de dispersión adimensional, y no está definido para el caso en que la media es cero. Un coeficiente de

variación mayor que uno indica la presencia de algunos valores erráticos en la muestra, los cuales pueden tener una gran influencia en la estimación. El coeficiente de variación está dado por la siguiente fórmula:

$$CV = \frac{\sigma}{\bar{X}}$$

### Coeficiente de Curtosis

Es una medida de cuán “afilado” es el pico de la distribución de los datos. Tradicionalmente el valor de este coeficiente se compara con cero (valor para el caso de una distribución normal). Un valor mayor que cero indica una distribución más puntiaguda que la normal y un valor menor que cero refleja una distribución más achatada que la normal. Sin un número de muestras grandes, el uso de este coeficiente es de dudoso valor (**Cruz, 2009**). El coeficiente de Curtosis está dado por la siguiente fórmula:

$$CC = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{\sigma^4}$$

Cuando se quiere conocer si los valores de una distribución están concentrados cerca de la media presentando un pico elevado, o si la forma de distribución es plana estando los valores más dispersos alrededor de la media, se hace uso del coeficiente de Curtosis.

### Estadígrafos que no serán implementados

**Media aritmética ponderada:** tiene en cuenta la importancia relativa de cada uno de los datos. No es calculado debido a que los datos de las muestras de los pozos de perforación poseen similar importancia.

**Moda:** es el valor más común. Una información puede tener una moda y se llama unimodal, dos modas y se llama bimodal, o varias modas y llamarse multimodal. Sin embargo puede ocurrir que la información no posea moda (**Salazar, 2002**). Por estas razones resulta difícil interpretar un conjunto de datos, descartándolo de la solución.

**Media armónica:** en caso de existir valores iguales a cero entre los datos geológico - mineros de los pozos de perforación, la media armónica no se puede determinar; además, no es aconsejable utilizarla cuando se trabaja con valores pequeños (**Martín Pilego, y otros, 2007**).

**Percentiles, Quintiles, Deciles:** dividen la información en cien, cinco y diez partes iguales respectivamente, agrupándolas en porcentajes de 1, 2, ..., 98, 99; 20, 40, 60, 80; y 10, 20, 30, 40, 50, 60,

70, 80 y 90 por ciento. No son calculados debido a que es calculada la medida de resumen Cuartiles, que permite a su vez determinar el Rango intercuartil, eliminando de esta forma los valores extremos de la distribución y midiendo la variabilidad de la mitad central de los datos.

## **Gráficos más utilizados para la representación de la distribución de los datos geológico – mineros**

Según el Dr José Manuel Becerra Espinosa el estudio estadístico permite encontrar el tipo de distribución de frecuencia que siguen los datos permitiendo una visualización rápida mediante gráficas sencillas (importante contribución para la interpretación y control de los futuros resultados) consiguiendo que un simple análisis visual ofrezca la mayor información posible. La gráfica posibilita la representación de los datos (generalmente numéricos) mediante líneas, superficies o símbolos, para ver la relación que guardan entre sí y sirve para analizar el comportamiento de un proceso, o un conjunto de elementos o signos que permitan la interpretación de un fenómeno.

### **Distribuciones**

Afirma el M.Sc Olman Ramírez Moreira: para que los datos sean útiles a la hora de distinguir patrones y tendencias, y llegar así a conclusiones lógicas, es necesario resumirlos y organizarlos. Una forma de organizar estos datos es clasificarlos por categorías o clases, y luego contar cuántas observaciones quedan dentro de cada categoría (datos agrupados). La importancia radica en que para el análisis e interpretación de los datos es necesario conocer la forma o patrón de su distribución, la posición de dicha distribución, es decir, alrededor de qué valor se tiende a concentrar los datos y la variabilidad de los mismos (dispersión de los datos alrededor de los valores centrales).

La distribución normal o gaussiana fue descubierta por Carl Gauss al estudiar el comportamiento de los procesos aleatorios. Es una función de densidad de probabilidad para la variable aleatoria continua y es ampliamente utilizada en la estadística y teoría de las probabilidades. Su representación es una curva simétrica parecida a una campana como se muestra en la *Ilustración 1*.

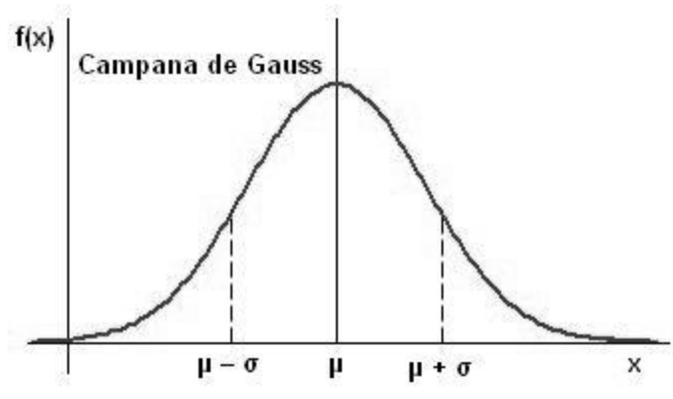


Ilustración 1 Distribución normal

Tiene como propiedades generales (**Alfonso Roche, 1989**):

- Es simétrica y tiene forma de campana.
- El área bajo de la curva es la probabilidad de ocurrencia de la variable.
- La mediana y la media son iguales.

La distribución normal o gaussiana es poco observado en los yacimientos minerales excepto en aquellos de origen sedimentario (**Cruz, 2009**).

En diferentes campos de aplicación se utilizan las distribuciones asimétricas con fines de análisis, ajustes, inferencia y predicción. Entre estas tiene especial interés la distribución logarítmica – normal, por la sencillez y facilidad de interpretación del esquema aleatorio que puede originar y por las propiedades que la caracterizan. Es una distribución de probabilidad de cualquier variable aleatoria que su logaritmo (neperiano, en base e) siga una distribución normal y se presenta con frecuencia en el estudio de histogramas asociados con leyes de muestras provenientes de yacimientos mineros (**Sironvalle, 2000**). Su representación es una curva asimétrica parecida a una campana como se muestra en la *Ilustración 2*.

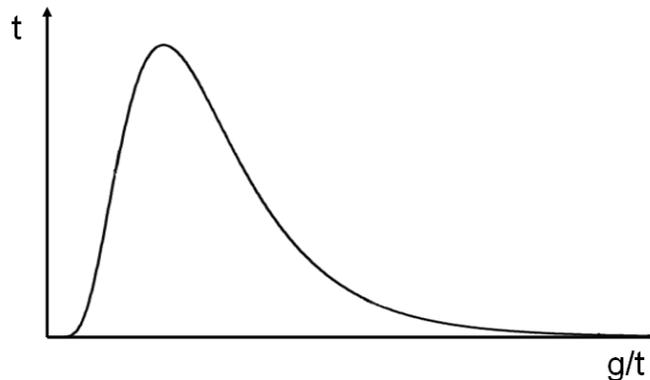


Ilustración 2 Distribución logarítmica – normal

Tiene como propiedades generales (**Alfonso Roche, 1989**):

- Presenta asimetría positiva.
- Puede ser transformada al modelo normal a partir del cambio de variables:  $L = \log(X)$
- Se cumple que: mediana < media.

Es común observar este tipo de distribución en problemas de evaluación de reservas donde existe una gran cantidad de valores bajos y unos pocos valores altos que definen al yacimiento (**Cruz, 2009**).

### Métodos gráficos para visualizar las características del conjunto de datos

El empleo de los gráficos para la visualización de la información ayuda a expresar ideas que se desean destacar. Son las representaciones visuales de los datos en donde se evidencia fundamentalmente la forma, la acumulación o tendencia, y la dispersión o variabilidad. Los gráficos no deben considerarse como sustitutos de un análisis estadístico, sino más bien como una ayuda visual del comportamiento de los datos.

Los **histogramas** son aplicables en todos los estudios en los que son necesarios analizar la pauta de comportamiento de un fenómeno en función de su frecuencia de aparición. Su utilización será beneficiosa para el desarrollo de todos aquellos proyectos abordados por equipos y grupos o individuos u organizaciones, implicados en la mejora de la calidad de los parámetros estudiados. Son un resumen gráfico de los valores producidos por las variaciones de una determinada característica de una variable

regionalizada<sup>4</sup>, representando la frecuencia con que se presentan. Permite resumir grandes cantidades de datos, analizarlos y evidenciar los esquemas de comportamiento y pautas de variación (elaborando una explicación admisible para dicha pauta, que relacione la variación con el fenómeno en estudio), transmitiendo la información de forma clara y sencilla (**FUNDIBEQ, 2010**).

Vale resaltar que la experiencia y habilidad del grupo en la interpretación y utilización de un histograma son fundamentales y que se debe profundizar en el conocimiento del proceso en estudio para utilizarlo de forma eficaz.

El **diagrama de caja** es una representación visual que describe simultáneamente varias características relevantes en un conjunto de datos, como por ejemplo el centro, la dispersión, la asimetría o simetría y la identificación de valores atípicos, es decir, observaciones que están desafortunadamente lejos del cuerpo principal de los datos. Debido a que incluso un solo punto inusual puede afectar de manera drástica el valor de algunas medidas numéricas (como la media y la desviación estándar), un diagrama de bloque está basado en medidas que son resistentes a la presencia de unos cuantos puntos inusuales. Este representa en un rectángulo (zona donde se halla el 50 % de los datos que particularmente se encuentran en el centro de la distribución) los tres cuartiles, la mediana y los valores mínimo y máximos. (**Johnson, y otros, 2003**).

La **tabla de frecuencia** consiste en el ordenamiento de la información obtenida de una muestra, durante el estudio de una variable. Es de gran utilidad cuando los datos estadísticos de que se disponen son numerosos, organizándolos en una tabla de manera tal que faciliten el trabajo. Agrupar los datos por intervalos en una tabla de frecuencias surge como una necesidad de estudiar variables continuas o discretas, donde el número de datos distintos se hace casi tan grande como el número de datos de la muestra.

El número de intervalos de la tabla de frecuencia estará dado según la fórmula de Sturges, determinado por la ecuación:  $m \cong 1 + 3.22 * \log_{10} n$ , donde  $n$  es el número de datos.

---

<sup>4</sup> Es aquella que el valor dado en un punto dentro del yacimiento está en función de su magnitud y de su soporte

## **Análisis de soluciones existentes**

MINITAB es un software diseñado para ejecutar funciones estadísticas básicas y avanzadas. Su ventaja fundamental radica en que combina lo amigable del uso de Microsoft Excel con la capacidad de ejecución del análisis estadístico. Realiza la estadística descriptiva y permite formular una gran variedad de gráficos. Este software fue desarrollado en la Universidad Estatal de Pennsylvania en el año 1972. Sin embargo la versión completa comercializada en el 2006 es de carácter privativo, por lo que la adquisición de este software traería consigo el oneroso pago de versiones, licencias y capacitación. Además, está sólo disponible para sistemas operativos Windows y es necesario tener una conexión a internet para la activación de las licencias (**Minitab Inc, 2013**). Por estas razones se descarta como posible solución al problema a resolver.

GEMCOM SURPAC es el software para la geología y planificación minera más utilizado en el mundo actualmente. Genera eficiencia y exactitud gracias a su facilidad de uso. Este programa abarca todos los requerimientos de los geólogos y es lo suficientemente flexible como para ser utilizado en cualquier cuerpo mineralizado. Brinda servicios de geología, posibilitando una mejor organización de los datos. Entre estos se encuentra el de análisis estadístico que tiene sus fundamentos en las estadísticas básicas (**Gemcom Software International Inc, 2013**). Sin embargo la herramienta que brinda para el análisis estadístico básico de datos sólo calcula los estadígrafos: mínimo, máximo y media aritmética (**Gemcom Software International Inc, 2013**). Además, por concepto del oneroso pago de versiones, licencias, capacitación y servicios, es descartado como posible solución al problema a resolver.

XLSTAT es la solución líder de estadística y análisis de datos para Microsoft Excel, ofreciendo una amplia variedad de funciones para mejorar las ya existentes. Tiene como ventaja su compatibilidad con todas las versiones de Excel, desde la de 1997 hasta la de 2011, excepto la versión 2008 para Mac, y es compatible con sistemas operativos Windows 9x a Windows 7, así como con sistemas Mac con base PowerPC e Intel. Tiene como desventajas su carácter privativo (**Addinsoft, 2013**), además de que al simular Microsoft Excel en el sistema operativo Linux no se asegura un funcionamiento óptimo, dado que Excel no contará con todas las dependencias necesarias como lo haría con Windows. Por estas razones es descartado como posible solución al problema a resolver.

SPSS es un paquete estadístico para el análisis de datos en las ciencias sociales y económicas. Es un grupo de programas y subprogramas conectados de manera tal que funcionan en conjunto permitiendo aplicar a un mismo fichero de datos un número ilimitado de procedimientos estadísticos de forma sincronizada. Prepara los datos para el análisis de forma rápida y sencilla consiguiendo informes tabulares, gráficos y diagramas de las distribuciones y de las tendencias (**Pérez López, 2001**). Por ser SPSS un software privativo es descartado como posible solución.

Matlab es un software aplicativo que permite su utilización en diferentes áreas del conocimiento, además permite la posibilidad de utilizar ToolBox<sup>5</sup> especializados que facilitan el trabajo y aumentan la funcionalidad del programa, como el caso del ToolBox estadístico. El paquete estadístico de Matlab ha sido desarrollado para proveer ayuda a cualquier tipo de área, desde las finanzas hasta la ingeniería, con herramientas interactivas capaces de establecer análisis detallados de datos. Viene acompañado de una completa serie de funciones para desarrollar desde las más básicas aplicaciones estadísticas hasta un completo diseño y proceso de cualquier análisis estadístico (**Franco Lemus, y otros, 2005**). Es un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio, lo que se dificulta su uso para usuarios comunes. Por esta razón es descartado como posible solución del problema a resolver.

PSPP es un programa para el análisis estadístico de los datos y un reemplazo libre para el programa SPSS. Tiene como ventajas que la copia de PSPP no va a caducar o dejar de trabajar deliberadamente en el futuro; no tiene límites artificiales en el número de casos o variables que se pueden utilizar; permite realizar la estadística descriptiva; puede realizar sus análisis tan rápido como sea posible sin importar el tamaño de los datos de entrada; no hay derecho de licencia; trae indexado el manual de usuario; entre otras. Tiene como desventaja que no realiza el cálculo de algunos estadígrafos importantes como por ejemplo: la mediana, la media geométrica, los cuartiles, la desviación media, el coeficiente de asimetría y el coeficiente de Curtosis (**Free Software Foundation, 2013**). Por estas razones es descartado como posible solución del problema a resolver.

RecMin como herramienta minera se ha venido desarrollando bajo la tutela del Dr. Ing. de Minas César Castañón Fernández. El software constantemente se está actualizando y perfeccionando y ofreciéndose de forma gratuita a la comunidad minera del mundo. Además, requiere de condiciones mínimas de

---

<sup>5</sup> Caja de herramientas.

hardware y de software para su funcionamiento y, está disponible en el idioma español (**SolMine, 2013**). Tiene como desventaja que el programa no realiza un análisis estadístico de los datos geológico - mineros, sólo los prepara para poder hacerlo en otros programas especializados. Por esta razón es descartado como posible solución.

R es un proyecto GNU desarrollado en los laboratorios Bell. Es un lenguaje y un entorno para computación y gráficos estadísticos. Está disponible como software libre bajo los términos de la Licencia de GNU Free Software Foundation. Además se compila y ejecuta en una amplia variedad de plataformas UNIX y sistemas similares (incluyendo FreeBSD y Linux), Windows y MacOS. Es un conjunto integrado de servicios de software para la manipulación, cálculo y representación gráfica de datos que incluye: un manejo eficaz de los datos y del almacenamiento, una gran colección coherente e integrada de herramientas intermedias para el análisis de los datos, facilidades gráficas para el análisis y visualización de los mismos y un bien desarrollado lenguaje de programación sencillo y eficaz (**The R Project for Statistical Computing, 1997**). Ejecuta los cálculos a través de comandos y viene integrado en el repositorio de Ubuntu. Tiene como desventajas, que resulta complejo vincularlo a una aplicación en C++ y por lo tanto es engorroso el proceso de compilación. Por esta razón es descartado como posible solución.

## **Metodología y tecnologías a utilizar en el desarrollo de la herramienta para el análisis exploratorio de los datos geológico – mineros de los pozos de perforación**

### **Metodología de desarrollo de software**

Las metodologías de desarrollo de software surgen como una necesidad de utilizar por parte de los desarrolladores “un conjunto de procedimientos, técnicas, herramientas y soporte documental para el desarrollo del software” (**Medina Patón, 2006**).

Se conocen dos clasificaciones de metodologías de desarrollo de software: las llamadas pesadas y las llamadas ágiles. Los desarrolladores en la actualidad se enfrentan al dilema de cuál de ellas escoger para dar solución a lo solicitado por el cliente. Para ello cada una propone roles que el grupo de desarrollo debe cubrir, realizar las actividades que les corresponden y generar los artefactos de cada tarea con su

documentación correspondiente. Actualmente no existe una metodología global que encierre características que puedan aplicarse a cualquier tipo de proyecto.

Características de las metodologías pesadas:

- Centradas en la definición detallada de los procesos y tareas a realizar. Requieren una extensa documentación.
- Son más eficaces y necesarias cuanto mayor sea el proyecto con respecto a tiempo y recursos.
- Requieren de una organización extrema.

Rational Unified Process (RUP) como metodología pesada es un “proceso de desarrollo de software que define *quién* está haciendo *qué*, *cuándo*, y *cómo* alcanzar un determinado objetivo”, además “es la definición del conjunto de actividades que guíen los esfuerzos de las personas implicadas en el proyecto, a modo de plantilla que explica los pasos para terminar el proyecto” (**Jacobson, y otros, 2004**). Tiene como características fundamentales el ser iterativo e incremental, dirigido por casos de usos y centrado en la arquitectura (**Pressman, 2007**). Sin embargo presenta un gran número de roles y de artefactos, y en la mayoría de los casos en el momento de su implantación se considera un proceso demasiado costoso, por lo que se descarta como metodología a utilizar.

Después de analizar las características de las metodologías pesadas, especialmente a la metodología RUP, se considera más ventajoso controlar el desarrollo de la herramienta para el análisis estadístico básico de los datos geológico - mineros de los pozos de perforación, mediante el uso de una metodología ágil, proporcionando un plano en que apoyarse el equipo de trabajo.

Características de las metodologías ágiles (**Villar, 2007**):

- Le da más importancia a crear un producto software que funcione que escribir mucha documentación, puesto que si el software no funciona la documentación no vale de nada. A nivel interno es necesario la documentación imprescindible, y a nivel externo lo que el cliente requiera.
- El cliente interactúa constantemente con el proyecto, ya que supone que la satisfacción con el producto será mayor.

- Es más flexible con respecto a los cambios para evitar el fracaso. La estrategia de planificación se basa en: planes detallados para las próximas semanas, planes aproximados para los próximos meses y planes generales para plazos mayores.
- Valora más al individuo y a las iteraciones del equipo que a las herramientas y procesos utilizados.

**Open Unified Process (OpenUP)** es una plataforma de proceso de desarrollo de software que cubre un conjunto amplio de necesidades y que toma un acercamiento al desarrollo de software ágil. Mantiene características esenciales de RUP como desarrollo iterativo e incremental, dirigido por casos de uso y centrada en la arquitectura. OpenUP es un proceso iterativo de desarrollo de software mínimo, pues sólo incluye el contenido fundamental; completo, ya que puede ser manifestado como un proceso entero para construir un sistema; y extensible, porque puede ser usado para añadir o ajustar el contenido del proceso (Iglesias, 2007).

Por tener conocimientos básicos de la metodología RUP y ser OpenUP un proceso con características similares, con la diferencia de que es un proceso mínimo y suficiente; contener un conjunto simplificado de roles, actividades, artefactos y guías; ser un proceso anti - burocrático y agnóstico en cuanto a herramientas (IDE, lenguajes, sistemas operativos, etcétera) que pueden ser usadas; permitir añadir tantos artefactos propios como se consideren necesarios, o ignorar los predefinidos y utilizar cualquier tipo de artefactos definidos por el equipo de desarrollo (siempre y cuando estos cumplan el objetivo para el cual fueron creados: almacenar información); y tener un desarrollo continuo para obtener realimentación y realizar las mejoras respectivas, es seleccionada OpenUP como metodología de desarrollo.

### **Lenguaje de Modelado**

#### **Lenguaje Unificado de Modelado (UML V2.0)**

Es el lenguaje de modelado de sistemas de software que proporciona la tecnología necesaria para apoyar la práctica de la ingeniería del software orientada a objetos. Es un lenguaje gráfico para visualizar, especificar, construir y documentar todos los elementos que forman un sistema de software, con una notación robusta para el modelado. Modelar con UML trae una serie de beneficios a los programadores, pues ofrece un estándar para describir un plano del sistema, incluyendo aspectos conceptuales como procesos de negocios y funciones del sistema, y otros concretos como expresiones de lenguajes de programación y componentes de software reutilizables, todo esto previo a la escritura del código. Aunque

UML es un lenguaje, este posee más características visuales que programáticas, facilitando así el entendimiento común entre los miembros de un equipo de trabajo, dígase analistas, diseñadores, programadores, etcétera. La aplicación del UML puede producir un arreglo de productos de trabajo (por ejemplo, modelos y documentos), sin embargo, éstos los reducen los ingenieros de software para lograr que el desarrollo sea más ágil y reactivo ante el cambio **(Pressman, 2000)**.

UML proporciona la capacidad de modelar actividades de planificación de proyectos y de sus versiones, expresar requisitos y las pruebas sobre el sistema, representar todos sus detalles así como la propia arquitectura. Mediante estas capacidades se obtiene una documentación que es válida durante todo el ciclo de vida de un proyecto **(Alarcón, 2000)**.

### **Herramienta CASE**

#### **Visual Paradigm V8.0**

Es una herramienta de Ingeniería de Software Asistida por Computación (CASE) que propicia un conjunto de ayudas para el desarrollo de programas informáticos, desde la planificación, pasando por el análisis y el diseño, hasta la generación del código fuente de los programas y su documentación correspondiente. Ha sido concebida para soportar el ciclo de vida completo del proceso de desarrollo del software (análisis y diseño Orientado a Objetos, construcción, pruebas y despliegue) a través de la representación de todo tipo de diagramas. Soporta los principales estándares de modelado como el UML y facilita la interoperabilidad con otras herramientas CASE y la mayoría de los entornos de desarrollo integrado.

Se caracteriza por:

- Multiplataforma.
- Licencia gratuita y comercial.
- Fácil de instalar y actualizar.
- Compatibilidad entre ediciones.
- Uso de un lenguaje estándar común a todo el equipo de desarrollo que facilita la comunicación **(Pressman, 2002)**.

## **Lenguaje de programación**

Un lenguaje de programación está conformado por una serie de reglas sintácticas y semánticas que son utilizadas por el programador a través de las cuales se creará un programa o subprograma. Todas estas instrucciones escritas por el desarrollador y que forman dicho programa es a lo que se denomina código fuente (**Martell, 2009**).

**C++** es un lenguaje de programación híbrido diseñado por Bjarne Stroustrup a mediados de los años 80 con la intención de extender el lenguaje C con mecanismos que permitan la manipulación de objetos. Es un lenguaje de propósito general, donde pueden implementarse desde sistemas operativos y compiladores hasta aplicaciones de bases de datos y procesadores de texto, pasando por juegos, aplicaciones a medida, entre otras.

Algunas de las características más importantes de este lenguaje se resumen a continuación (**Amaya, 2013**):

- Compila directamente al código nativo de la máquina, lo que le permite ser uno de los más rápidos idiomas en el mundo.
- Ofrece un apoyo notable para los paradigmas de programación (procedimental, estructurada, genérica y orientada a objetos).
- Como uno de los idiomas más utilizados en el mundo y como un lenguaje abierto, tiene un amplio rango de compiladores que se ejecutan en diferentes plataformas.

## **Herramienta del Entorno de Desarrollo Integrado**

### **QtCreator V2.5.0**

QtCreator es un entorno de desarrollo integrado para aplicaciones multiplataforma que se utiliza ampliamente para el desarrollo de software con interfaz gráfica de usuario (GUI). Se centra en proporcionar características que ayudan a los nuevos usuarios de Qt (resaltado de sintaxis, la finalización de código, apoyo a la refactorización del código fuente, entre otras) y también en aumentar la productividad de los desarrolladores con experiencia. Integrado con QtDesigner y el constructor de formularios para C ++, lo que permite diseñar y construir rápidamente Widgets y diálogos que se utilizarán en la aplicación. Los formularios están totalmente funcionales, y se pueden pre visualizar inmediatamente

para asegurar lo que realmente se desea. Permite al usuario generar un proyecto para una aplicación de consola, una aplicación GUI, o una librería de C++ (**Qt Project Hosting, 2011**).

Algunas de las características más importantes de este IDE son (**Meyer, 2007**):

- Compatibilidad multiplataforma.
- Performance de C++.
- Disponibilidad del código fuente.
- Excelente documentación.
- Fácilmente internacionalizable.

## **Sistema Gestor de Base de Datos**

### **PostgreSQL 9.1**

Un Sistema Gestor de Base de Datos (SGBD) es una colección de programas que permiten al usuario gestionar una base de datos, manteniendo consistente el conjunto de datos almacenados. Un SGBD relacional es aquel donde los datos almacenados cumplen una serie de normas y reglas y son percibidos por los usuarios como un conjunto de tablas interrelacionadas (**Garzón Pérez, 2010**).

PostgreSQL es un sistema gestor de base de datos objeto - relacional debido a que incluye características como la herencia, tipos de datos, funciones, restricciones, entre otras, pero no puede ser considerado un SGBD puramente orientado a objetos. Utiliza el modelo Cliente - Servidor y es considerado el SGBD libre más avanzado soportando la gran mayoría de las transacciones SQL<sup>6</sup>. Tiene la ventaja de que puede ejecutarse en la mayoría de los sistemas operativos más utilizados a nivel mundial (**Guerrero, 2010**).

Está licenciado bajo BSD<sup>7</sup> por lo que se tiene libertad de usar, modificar y distribuir en productos comerciales o no comerciales de forma gratuita. Por estas razones es utilizado como SGBD.

### **Biblioteca numérica**

ALGLIB es una biblioteca numérica de análisis y procesamiento de datos compatible con varios de los lenguajes de programación más utilizados, entre los cuales se encuentran C++, C#, Pascal, VBA y, los

<sup>6</sup> Lenguaje de Consulta Estructurado

<sup>7</sup> Distribución de Software Berkeley

sistemas operativos Windows, Linux, Solaris. Además, es de código abierto por lo que se tiene libertad de usar de forma gratuita bajo la licencia GPL (Licencia Pública General de GNU). Entre las características que incluye se encuentra la estadística descriptiva. ALGLIB tiene como desventaja que sólo permite calcular la media aritmética y la mediana entre los estadígrafos de tendencia central, la desviación estándar y la desviación media entre los estadígrafos de dispersión y, el coeficiente de asimetría, el coeficiente de Curtosis y los percentiles entre los estadígrafos de forma. Por esta razón es descartada como biblioteca numérica a utilizar (**ALGLIB Project, 2013**).

Según (**Free Software Foundation, 2012**) **GSL** es una biblioteca numérica para programadores de C++ que ofrece una amplia gama de rutinas matemáticas, en donde, entre los temas tratados se incluye la estadística. Entre los beneficios que brinda esta biblioteca se encuentran:

- Es de libre acceso para todos.
- No impone ninguna condición interna para su uso.

Tiene como características principales:

- Está dirigido a usuarios ordinarios o científicos. Cualquiera con conocimientos básicos de programación de C++ será capaz de utilizar la biblioteca.
- GSL se distribuye bajo los términos de la licencia GPL.

El uso de GSL en la herramienta para el análisis estadístico de los datos geológico – mineros está dado por ser el punto de partida en la implementación de los estadígrafos: media aritmética, media geométrica, mediana, mínimo y máximo, coeficiente de asimetría, coeficiente de Curtosis, desviación estándar, cuartiles, varianza y la desviación absoluta, permitiendo a la vez la implementación de otros estadígrafos como por ejemplo: el rango, el rango intercuartil, el centro de amplitud y el coeficiente de variación. El nombre de los métodos y de las variables que son pasadas por parámetros, y las fórmulas que sirvieron como referencia para implementar los estadígrafos, fueron tomadas de dicha biblioteca, con el objetivo de brindar la bibliografía a consultar para posteriores estudios.

Además, con el propósito de evitar la vinculación de bibliotecas externas con la aplicación (la cual debe ser multiplataforma), y de esta manera también evitar el esfuerzo por lograr que funcionen para todas las plataformas (en el caso de que se fuera a montar el sistema en un centro de trabajo con sistema operativo

Linux habría que instalar cada librería externa usada) se decidió implementar los estadígrafos en lenguaje C++. Por último y no menos importante, se piensa en un futuro la posible adición de nuevas fórmulas para realizar el análisis estadístico.

### **Biblioteca de graficación**

Para apoyar el análisis estadístico se emplearon los gráficos (ayudando a la visualización de la información) generados por la biblioteca **QCustomPlot. 1.0.0**. Esta brinda una variedad de gráficos y la posibilidad de graficar los datos de manera sencilla, permitiendo identificar características de las variables en estudio como por ejemplo: concentración, valores atípicos, dispersión, entre otras. Al igual que la biblioteca numérica GSL, QCustomPlot se encuentra dentro de la aplicación como una clase más, para evitar de esta forma los mismos agravios.

## **Conclusiones parciales**

- El análisis exploratorio de los datos geológico – mineros de los pozos de perforación permite definir características de los minerales tales como valores atípicos, concentración y forma de distribución dentro del yacimiento, y presentarlas de una forma simple y clara, ayudando al analista de datos a la toma de decisiones que guíen las operaciones en las empresas y compañías geomíneras.
- Los software mineros y estadísticos imperantes en el mundo, no satisfacen los requerimientos específicos de la industria minera nacional como por ejemplo: software libre, multiplataforma, adaptado a las necesidades, entre otras. Además, resulta lento e impreciso el cálculo de los estadígrafos y la representación de la información a través de gráficos debido a que se maneja un número considerable de datos geológico – mineros. Por esto se decide llevar a cabo el desarrollo de la herramienta para el análisis exploratorio de los datos geológico – mineros de los pozos de perforación.
- En el país son usados software de minería adquiridos en el mercado internacional. Un problema existente en la actualidad es la compra de estos programas por parte de empresas norteamericanas, lo que no permite que el país actualice los nuevos cambios producidos en los mismos. Además, las herramientas informáticas de procedencia nacional que están actualmente en explotación no resuelven todos los problemas a los que se enfrenta la industria minera, por lo que se aboga al desarrollo de un sistema que integre todas las actividades relativas al minado.
- La correcta selección de la metodología de desarrollo de software y de las tecnologías y herramientas a ser usadas en el desarrollo de un sistema, facilita en gran medida el trabajo para dar solución al problema planteado. La selección de OpenUP como metodología de desarrollo de software se fundamenta en que es un proceso de desarrollo con un acercamiento ágil y que genera la documentación necesaria para el entendimiento de la solución propuesta. Por último, la selección de las tecnologías y herramientas libres tiene su base en la política de independencia tecnológica que sigue el país.

## Capítulo 2: Características y diseño del sistema

Para el desarrollo de un software de cualquier tipo, el lenguaje que deben dominar todos los participantes debe ser uno común, para que los especialistas realicen el proceso de ingeniería de forma fácil y sencilla, y para compartir el conocimiento con los involucrados en el desarrollo. Este lenguaje común se desprende de la interpretación del modelado del dominio, el cual contribuye a la comprensión del contexto del problema y de los requisitos del sistema que se derivan de este contexto.

Cuando el grupo de desarrolladores de software se enfrenta al desarrollo de un sistema, buscan la forma de extraerles a los clientes lo que ellos necesitan. Este acto de descubrir es lo que se conoce como definición de requisitos, generalmente en circunstancias difíciles. Además, el esfuerzo principal durante la definición de los requisitos recae sobre el desarrollo de un modelo del sistema, siendo la utilización de los casos de uso una forma adecuada de crear este modelo. A este modelo del sistema se le encuentra su forma, incluida la arquitectura, durante el diseño.

En el presente capítulo se realiza la propuesta del modelo del dominio y del sistema correspondiente a la herramienta para el análisis estadístico básico de los datos geológico - mineros de los pozos de perforación. Se definen las clases y el diagrama del modelo del dominio, los requisitos funcionales y no funcionales que debe cumplir la aplicación a desarrollar, los actores sistema, el diagrama de casos de uso del sistema y las descripciones de los mismos. Por último, se define el diagrama de clases del diseño y la descripción de cada clase.

### Modelado del Dominio

El modelado del dominio captura los tipos de objetos o clases más importantes en el contexto del sistema, mostrando a los clientes, desarrolladores, usuarios y revisores cómo se relacionan unas con otras mediante asociaciones. El resto de los objetos o clases candidatas que pueden ser extraídas del contexto del dominio se guardan como definiciones en un glosario de términos. El objetivo del modelo del dominio radica en utilizar un vocabulario común entre usuarios, clientes, desarrolladores y otros interesados (Jacobson, y otros, 2004).

### Descripción de las clases del Modelo del Dominio

- **Análisis estadístico:** técnicas estadísticas que tienen como objetivo examinar los datos de los pozos de perforación, obteniendo un entendimiento básico de ellos y de las características de la variable analizada.
- **Barra:** utilizado para variable de tipo discreta y que permite representar la frecuencia que pertenece a una categoría en particular.
- **Caja:** utilizado para representar conjuntos de datos e identificar características tales como distribución, concentración y presencia de valores atípicos.
- **Datos:** valores de las muestras obtenidas del suelo durante las campañas de exploración.
- **Dispersión:** se relaciona con la mayor o menor concentración de los datos en torno a un valor central, generalmente la media aritmética.
- **Estadígrafos:** valores numéricos obtenidos a través de cálculos, que caracterizan a las muestras (obtenidas de los pozos de perforación) en estudio.
- **Forma:** comunican información sobre la forma de una distribución de datos como por ejemplo: simetría, longitud de las colas, etcétera.
- **Gráficas:** representación de los datos geológico - mineros de los pozos de perforación mediante el uso de recursos gráficos como por ejemplo: líneas, vectores, superficies, símbolos, puntos plasmados en coordenadas cartesianas, entre otros.
- **Minerales:** sustancias naturales generalmente de composición química, homogéneas y de origen inorgánico. Pueden hallarse en la superficie de la tierra formados de manera espontánea y, resultan una parte útil de una explotación minera.
- **Muestras por intervalos:** es una parte ó porción de subsuelo extraída de un conjunto que permiten considerarla como representativa del mismo. Permiten identificar las características de los minerales.
- **Pozos de perforación:** es el resultado de la utilización de un método de perforación, que posibilita la obtención de muestras de roca para su posterior análisis.

- **Posición:** describe a través de un sólo número, dónde están localizadas las distintas partes de una distribución.
- **Tabla de frecuencias:** es la ordenación de los datos obtenidos de una muestra, en el estudio de una sola variable. Está formada por las categorías y sus frecuencias correspondientes.
- **Yacimiento mineral:** concentración selectiva de minerales, muy por encima de sus valores normales para un tipo determinado de roca, dando origen a concentraciones irregulares.

**Diagrama de clases del Modelo del Dominio**

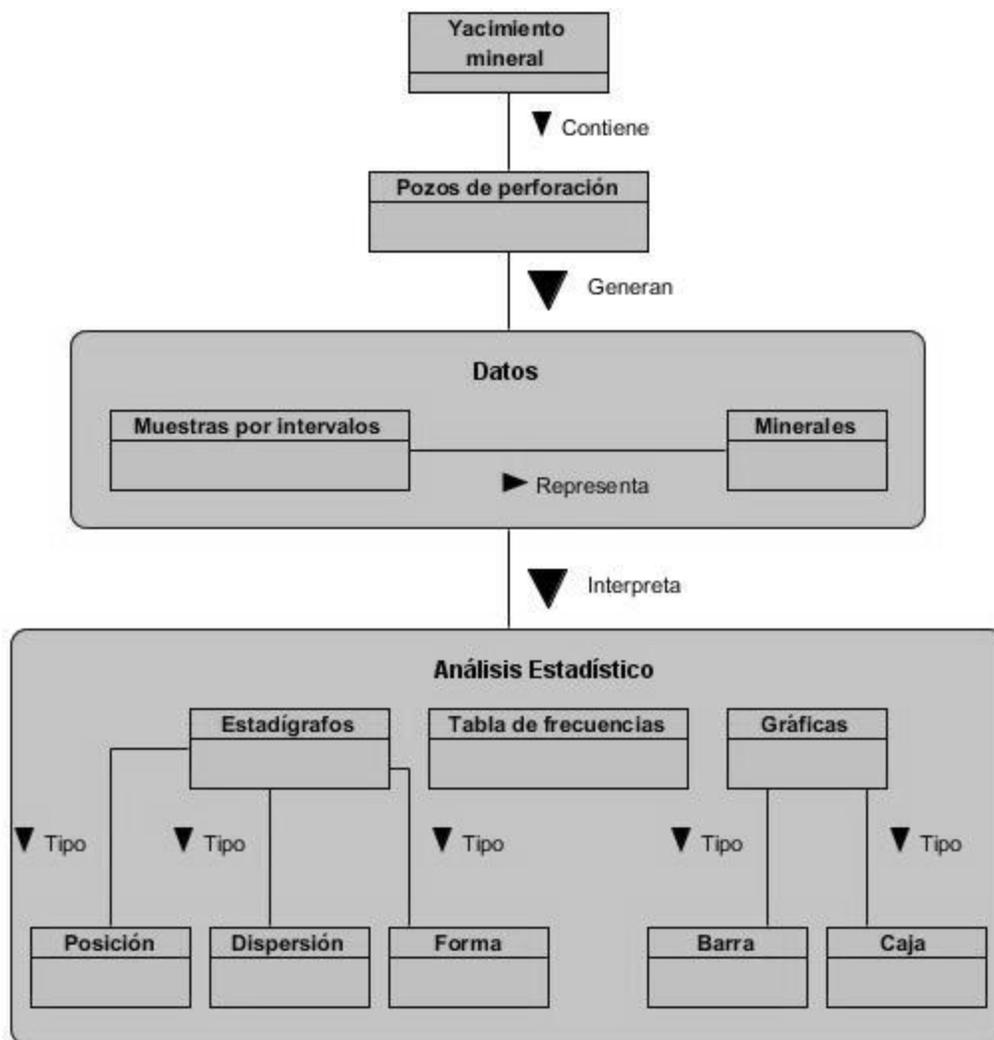


Ilustración 3 Diagrama de clases del Modelo del Dominio

### Descripción del diagrama de clases del Modelo del Dominio

Un yacimiento mineral, en este caso nacional, se explora mediante varios pozos de perforación. Estos a su vez se fragmentan en varias muestras por intervalos que representan a los minerales y brindan datos geológico – mineros de los yacimientos. Dichos datos permiten definir las características de los minerales a través de un análisis estadístico realizado sobre los mismos mediante el uso de estadígrafos (posición, forma y dispersión), tabla de frecuencias y gráficas (barra y caja). Con el análisis exploratorio de los datos geológico - mineros se obtiene la información que determina las características de la variable en estudio.

### Requisitos

El propósito fundamental del levantamiento de los requisitos es la confección de una guía lo suficientemente buena para el desarrollo del sistema correcto. En esta situación es necesario un firme conocimiento del contexto en que se emplaza el software. Esto se logra con una descripción de los requisitos, con lo cual se llega a un acuerdo entre el usuario final y los desarrolladores sobre lo que el sistema debe o no debe hacer. Además se conoce que el cliente en la mayor parte de las veces será un especialista no informático, que debe ser capaz de leer y comprender el resultado de la captura de los requisitos. Por lo tanto, esta captura debe realizarse utilizando el lenguaje del cliente y, con cautela a la hora de introducir en los resultados formalidad, estructura y detalles sobre el funcionamiento interno del software. Por último, el resultado del levantamiento de los requisitos también ayuda al jefe de proyecto a planificar las iteraciones y las versiones que se le irán mostrando a los receptores (**Jacobson, y otros, 2004**).

### Requisitos Funcionales

Los requisitos funcionales del sistema son las condiciones o capacidades que el sistema debe cumplir (**Sommerville, 2005**).

A continuación se definen los requisitos funcionales que debe cumplir el sistema que se propone:

- **RF 1** Conectar a la base de datos: esta funcionalidad debe permitir conectarse a la base de datos para obtener los datos geológico - mineros de los pozos de perforación.
- **RF 2** Calcular la media aritmética de los datos (sin agrupar y agrupados por intervalos) geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.

- **RF 3** Calcular la media geométrica de los datos (sin agrupar y agrupados por intervalos) geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 4** Calcular la mediana de los datos (sin agrupar y agrupados por intervalos) geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 5** Calcular el mínimo de los datos geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 6** Calcular el máximo de los datos geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 7** Calcular los cuartiles de los datos (sin agrupar y agrupados por intervalos) geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 8** Calcular la desviación estándar de los datos (sin agrupar y agrupados por intervalos) geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 9** Calcular la varianza de los datos (sin agrupar y agrupados por intervalos) geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 10** Calcular el coeficiente de variación de los datos (sin agrupar y agrupados por intervalos) geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 11** Calcular la desviación media de los datos (sin agrupar y agrupados por intervalos) geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 12** Calcular el rango intercuartil de los datos (sin agrupar y agrupados por intervalos) geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 13** Calcular el coeficiente de asimetría de los datos (sin agrupar y agrupados por intervalos) geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 14** Calcular el coeficiente de Curtosis de los datos (sin agrupar y agrupados por intervalos) geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 15** Calcular el centro de amplitud de los datos geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.

- **RF 16** Calcular el número total de muestras tomadas de los pozos de perforación.
- **RF 17** Calcular el rango de los datos geológico - mineros de los pozos de perforación para la distribución normal y logarítmica - normal.
- **RF 18** Generar el histograma para los datos geológico - mineros de los pozos de perforación para la distribución normal.
- **RF 19** Generar el histograma acumulativo para los datos geológico - mineros de los pozos de perforación para la distribución normal.
- **RF 20** Generar la tabla de frecuencia para los datos geológico - mineros de los pozos de perforación para la distribución normal.
- **RF 21** Generar el gráfico de caja para los datos geológico - mineros de los pozos de perforación para la distribución normal.
- **RF 22** Generar un reporte con los resultados: esta funcionalidad debe permitir generar un reporte de los resultados de las medidas de resumen calculadas y la tabla de frecuencia obtenida en formato PDF.
- **RF 23** Exportar las imágenes generadas con los datos geológico - mineros de los pozos de perforación en formato PNG.

### **Requisitos No Funcionales**

Los requisitos no funcionales son fundamentalmente requisitos que no se pueden asociar a ningún caso de uso concreto, pero pueden tener impacto en varios casos de uso o en ninguno. Son aquellos que no se refieren directamente a las funciones específicas que proporciona el sistema, sino a las propiedades emergentes de éste como las restricciones arquitectónicas, seguridad, disponibilidad, usabilidad, entre otros (**Sommerville, 2005**).

A continuación se exponen los requisitos no funcionales que debe cumplir el sistema propuesto.

### **Requisitos de hardware:**

- Procesador Pentium 4 o superior.
- Memoria RAM 512 MB o superior.

**Requisitos de software:**

- Sistema operativo Windows XP o superior y/o distribuciones de Ubuntu 11.4 o superior.

**Requisitos de usabilidad:**

- La herramienta de análisis estadístico debe asegurar su uso para el analista de datos con conocimientos básicos en computación.
- La herramienta de análisis estadístico debe proporcionar una manipulación fácil de las funciones (calcular estadígrafos, graficar, etcétera) al analista de datos, con el objetivo de proveer aceptación en el uso de la misma.

**Requisitos de diseño:**

- Utilización de C++ como lenguaje de programación con vista a una posible integración con el sistema SYAM.
- Utilización de PostgreSQL como SGBD.
- Utilización de convenciones de nombrado y estándares de codificación.

**Requisitos de interfaz:**

- La aplicación debe tener una apariencia profesional y un diseño gráfico sencillo.
- El sistema debe permitir al usuario transitar de una tarea a otra sin necesidad de obligarlo a realizar acciones innecesarias o no deseadas.

**Requisitos legales, de derecho de autor y otros:**

- Como producto, se distribuye amparado bajo las normativas legales establecidas en el registro comercial emitido por las entidades jurídicas de la Universidad de las Ciencias Informáticas.
- Las tecnologías empleadas en el desarrollo de la herramienta deben ser tecnologías libres.

**Modelo del Sistema**

Es una representación gráfica que describe los procesos del negocio, el problema a solucionar y el software que tiene que ser desarrollado. Es una abstracción del sistema en estudio, simplificando y

resaltando de forma deliberada las características más relevantes. Es a menudo más comprensible que las descripciones detalladas en lenguaje natural de los requisitos del sistema (Sommerville, 2005).

### Actores del Sistema

Son los trabajadores del negocio que realizan las tareas a automatizar y, también puede ser algún actor del negocio siempre y cuando interactúe con el sistema (Jacobson, y otros, 2004). Después de haberse definido los requisitos funcionales y no funcionales que el sistema debe satisfacer, corresponde identificar quiénes son los actores del sistema.

Actor del Sistema	Justificación
Analista de datos	Es el encargado de calcular los estadígrafos, graficar los resultados y generar los reportes.

Tabla 2 Actor del Sistema

### Casos de Uso del Sistema

Un Caso de Uso del Sistema (CUS) es un fragmento de la funcionalidad que el sistema ofrece para aportar un resultado de valor para sus actores. Estos se forman agrupando requisitos funcionales y especifican las acciones que tienen lugar durante la interacción actor - sistema. Los CUS de la herramienta para el análisis estadístico de los datos geológico – mineros se definen a continuación:

- **CUS 1:** Capturar datos.
- **CUS 2:** Generar tabla de frecuencias.
- **CUS 3:** Calcular frecuencias.
- **CUS 4:** Generar gráficas.
- **CUS 5:** Calcular estadígrafos.
- **CUS 6:** Generar reportes.

**Diagrama de casos de uso del sistema**

Un Diagrama de Casos de Uso del Sistema (DCUS) representa gráficamente a los procesos y sus interacciones con los actores. En la siguiente ilustración se presenta el DCUS de la herramienta para el análisis estadístico de los datos geológico - mineros de los pozos de perforación.

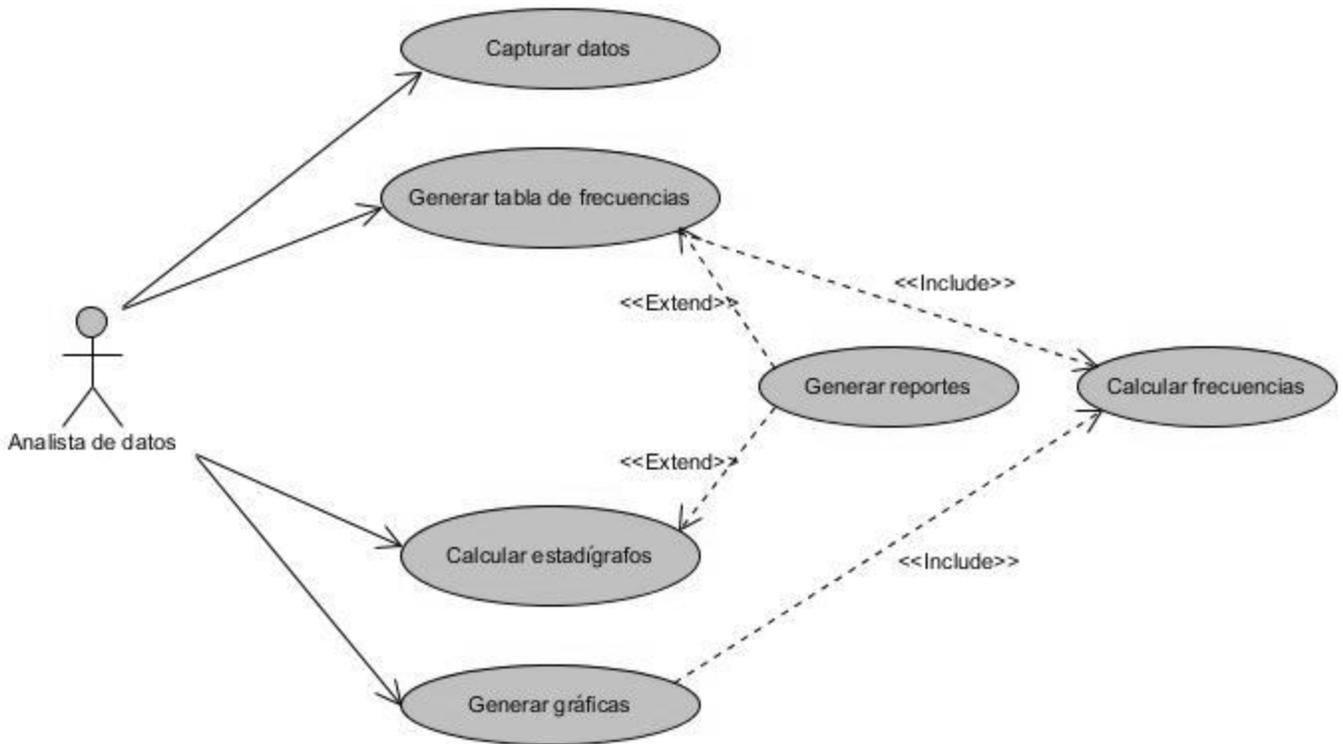


Ilustración 4 Diagrama de Casos de Uso del Sistema

**Descripción textual de los casos de uso del sistema**

Descripción del Caso de Uso del Sistema Calcular estadígrafos

<b>Caso de Uso</b>	Calcular estadígrafos.
<b>Actores</b>	Analista de datos: (Inicia) Selecciona obtener datos.
<b>Resumen</b>	Este caso de uso permite al analista de datos obtener los valores producidos por los estadígrafos una vez hecho los cálculos sobre los datos geológico - mineros.

<b>Precondiciones</b>	La conexión a la base de datos se ha establecido y se han obtenido los datos.	
<b>Referencias</b>	RF 2, RF 3, RF 4, RF 5, RF 6, RF 7, RF 8, RF 9, RF 10, RF 11, RF 12, RF 13, RF 14, RF 15, RF 16 y RF 17.	
<b>Prioridad</b>	Crítico.	
<b>Flujo de eventos</b>		
<b>Flujo básico "Calcular estadígrafos"</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Selecciona obtener datos.	
2.		Realiza los cálculos de los estadígrafos: media aritmética para datos sin agrupar y agrupados, media geométrica para datos sin agrupar y agrupados, mediana para datos sin agrupar y agrupados, mínimo, máximo, primer cuartil para datos sin agrupar y agrupados, segundo cuartil para datos sin agrupar y agrupados, tercer cuartil para datos sin agrupar y agrupados, desviación estándar para datos sin agrupar y agrupados, varianza para datos sin agrupar y agrupados, coeficiente de variación para datos sin agrupar y agrupados, coeficiente de Curtosis para datos sin agrupar y agrupados, coeficiente de asimetría para datos sin agrupar y agrupados, desviación media para datos sin agrupar y agrupados, rango, rango intercuartil para datos sin agrupar y agrupados, centro de amplitud y el total de muestras.

3.		Muestra los resultados.
4.		Termina el caso de uso.

<b>Relaciones</b>	<b>CUS extendidos</b>	Generar reporte: Paso 1 del Flujo Básico.
<b>Pos condiciones</b>		Se han calculado los estadígrafos y se han habilitado las funciones correspondientes a generar reporte y generar gráficas.

Tabla 3 Descripción del Caso de Uso del Sistema Calcular estadígrafo

Las descripciones textuales de los CUS Generar tabla de frecuencias, Calcular frecuencias, Generar gráficas, Capturar datos y Generar reportes, están recogidos en el **Anexo 1**.

## Modelo de Diseño

Para la transición de los requisitos al diseño normalmente se utiliza como intermediario el flujo de trabajo **Análisis**. Este proporciona una comprensión detallada de los requisitos mediante la refinación de los casos de uso y la asignación inicial de funcionalidad del sistema a un grupo de objetos que enfocan el comportamiento, funcionando como primera aproximación del modelo de diseño.

Es importante aclarar que el modelo de análisis hace abstracciones, evita resolver algunos problemas y tratar algunos requisitos que se considera que es mejor posponerlos para el diseño y la implementación. Esto trae consigo que en ocasiones no se pueda conservar la estructura proporcionada por el análisis, sino que se debe revisar y analizar durante el diseño y la implementación, debido a que el diseño debe considerar la plataforma de implementación, lenguajes de programación, sistemas operativos, entre otros **(Jacobson, y otros, 2004)**.

Como se explicó en el Capítulo 1 epígrafe "**Metodología y tecnologías a utilizar en el desarrollo de la herramienta para el análisis exploratorio de los datos geológico – mineros de los pozos de perforación**", la metodología OpenUP es un proceso configurable que puede ser adaptado de acuerdo a las especificaciones de cada proyecto, por lo que no se está obligado a hacer uso de todas las actividades y entregables definidos por la fundación Eclipse, sino que se puede configurar el proceso para que se adapte únicamente a aquellas partes que se consideran necesarias. Por lo tanto se decidió prescindir de la realización del modelo de análisis en el desarrollo de esta investigación debido a que:

- Se cuenta con requisitos simples, bien conocidos y con cierta comprensión de los mismos, por lo que no resulta difícil obtener una visión general del sistema.
- Variaría la estructura generada en el análisis dado que el diseño debe considerar las tecnologías a utilizar en la construcción del sistema.
- Se logra evitar el costo en tiempo y recursos de realizar las tareas de este flujo.

Teniendo en cuenta que se decidió no realizar el modelo de análisis, se puede pasar directamente al modelo de diseño. En el diseño se modela el sistema y define su forma, lo que incluye la arquitectura para

que soporte todos los requisitos funcionales y no funcionales. Además, el modelo de diseño sirve de abstracción de la implementación del sistema y es, de ese modo, utilizada como una entrada fundamental de las actividades de implementación. En el modelo de diseño, los casos de uso son realizados por las clases de diseño y sus objetos (**Jacobson, y otros, 2004**).

### Arquitectura de la herramienta

La arquitectura de software es la organización fundamental de un sistema encarnada en sus componentes, las relaciones entre ellos y el ambiente y los principios que orientan su diseño y evolución (**Billy Reynoso, 2004**). Para el desarrollo del presente trabajo se seleccionaron como estilos arquitecturales: Cliente-Servidor, en Capas y Orientado a Objetos.

El **Cliente-Servidor** es un estilo de las aplicaciones de software que presenta una máquina cliente que requiere de un servicio de una máquina servidor (no necesariamente en computadoras distintas), es decir, el cliente envía un mensaje solicitando un determinado servicio al servidor, y este envía uno o varios mensajes con la respuesta. Tiene como base fundamental separar el complejo software en módulos con la finalidad de hacer más fácil el desarrollo y mejorar su mantenimiento. Este estilo facilita la integración de nuevas tecnologías y la expansión de la infraestructura computacional, favoreciendo así la escalabilidad de las soluciones (**Jacobson, y otros, 2004**).

El estilo en **Capas** define cómo organizar el modelo de diseño en capas, de tal forma que los componentes de una capa sólo pueden hacer referencia a componentes en capas inmediatamente inferiores. La aplicación de este estilo trae como beneficio la simplificación y la organización del desarrollo de sistemas, reduciendo las dependencias y, que las capas más bajas no tengan conocimiento de ningún detalle de las superiores (**Jacobson, y otros, 2004**). Este estilo tiene muchas ventajas, entre las más importantes se encuentra: la separación entre las funcionalidades de cada capa, facilitando que si es necesario realizar algún cambio, sólo se tendría que realizar las correcciones necesarias en la capa requerida sin necesidad de revisar otros niveles.

El estilo **Orientado a Objetos** define al sistema como un conjunto de objetos que cooperan entre sí en vez de como un conjunto de procedimientos. Entre sus características se encuentran: describe el uso de objetos que contienen los datos y el comportamiento para trabajar con dichos datos, con un rol o responsabilidad distinta; fomenta la reutilización a través de la encapsulación y la herencia; difiere con el acercamiento procedimental en donde hay una secuencia de tareas y acciones predefinidas (el enfoque

orientado a objetos emplea el concepto de objetos interactuando uno con otros para realizar las tareas); entre otras (**de la Torre Llorente, y otros, 2010**). Este estilo tiene un considerable número de ventajas, entre las cuales una de las más importantes es: la descomposición de problemas en un conjunto de agentes en interacción para realizar las tareas.

Además, se utiliza el patrón **Modelo – Vista** al usar el componente de interfaz QTableView, el cual está especializado en la visualización de la información que recibe a través de los modelos. Este patrón se evidencia en el momento que se presenta un modelo donde es almacenado los datos y, una vista que es la encargada de mostrar dichos datos.

### **Patrones de diseño**

Los patrones de diseño constituyen la base fundamental para la búsqueda de soluciones a problemas comunes que surgen durante el desarrollo de un software. Estos, haciendo uso de métodos y estructuras que han sido probados en diversos sistemas, guían mediante expresiones y elementos el proceso de desarrollar cualquier aplicación informática. Se pueden utilizar como un conjunto estructural de relaciones entre clases, componentes y módulos para abstraer la posible implementación de la solución a desarrollar (**Larman, 1999**).

### **Patrones GRASP<sup>8</sup>**

Un sistema orientado a objetos está compuesto por objetos que se comunican mediante mensajes para llevar a cabo las operaciones. Por lo tanto, es imprescindible codificar buenos principios y sugerencias relacionados con la asignación de responsabilidades expresadas en forma de patrones, para tener calidad y eficiencia en el sistema software desarrollado. Algunos de estos patrones son:

- **Alta cohesión:** este patrón busca asignarle a cada clase la responsabilidad necesaria para su correcto funcionamiento, permitiendo que cada clase pueda colaborar con otras para realizar diversas operaciones. Esto proporciona que el software sea flexible a grandes cambios. Cada clase de la herramienta para el análisis estadístico de los datos geológico - mineros contiene responsabilidades estrechamente relacionadas y que representan el cumplimiento de los requisitos funcionales, como por ejemplo la clase `CalcularFrecuencia` contiene métodos que permiten el cumplimiento del requisito funcional generar la tabla de frecuencias.

---

<sup>8</sup> Patrones Generales de Software para Asignar Responsabilidades

- Experto: está diseñado para que la responsabilidad de realizar una operación recaiga sobre la clase que tiene la información necesaria. En la herramienta para el análisis estadístico de los datos geológico - mineros existen clases en donde se refleja dicho patrón, entre otras `CalcularFrecuencia`, `Reporte`, `ConectarseBD`, `AnalisisEstadistico` y `DatosAgrupados`, teniendo como resultado un diseño con alta cohesión y bajo acoplamiento. Como por ejemplo la clase `AnalisisEstadistico` contiene la información necesaria para crear un objeto de tipo `Reporte`, y esta a su vez utiliza dicha información para implementar el método `QTextDocument reporteDelAnalisisEstadistico()`.
- Creador: es el encargado de ayudar a identificar quién debe ser el responsable de la instanciación de nuevas clases, dado que la creación de objetos es una de las actividades más frecuentes en un sistema orientado a objetos. Basándose en reconocer si dicho creador tiene la información suficiente y necesaria para realizar dicha actividad, al escogerlo, se da soporte al bajo acoplamiento. En la herramienta para el análisis estadístico de los datos geológico - mineros existe la clase `AnalisisEstadistico`, en donde se refleja dicho patrón, teniendo como resultado la reusabilidad y el encapsulamiento.
- Bajo acoplamiento: su función es establecer las conexiones entre las clases sólo cuando son necesarias, obteniendo como beneficio al realizar alguna modificación a una clase en específico, la mínima repercusión en el resto de los objetos que puedan contener al modificado. En la herramienta para el análisis estadístico de los datos geológico - mineros existe un bajo acoplamiento expresado en la escaza dependencia entre clases. Como por ejemplo las clases `ConectarseBD`, `Reporte`, `DatosAgrupados` y `CalcularFrecuencia` no dependen de la presencia de otras clases para cumplir con su objetivo, facilitando así su reutilización.
- Controlador: encargado de asignar la responsabilidad de controlar el flujo de eventos del sistema a clases específicas. Es utilizado en la herramienta para el análisis estadístico de los datos geológico - mineros ya que se cuenta con una clase controladora (`AnalisisEstadistico` recibe datos y los envía a las distintas clases según la funcionalidad requerida) encargada de interactuar con las restantes clases del sistema.

### Patrones GoF<sup>9</sup>

Representan un vocabulario de principios y métodos susceptibles de combinarse para diseñar objetos, asignando responsabilidades a estos y diseñando la colaboración entre ellos. Consisten en soluciones concretas y técnicas basadas en la programación orientada a objetos, utilizadas frecuentemente debido a que están basadas en la experiencia obtenida al resolver problemas reiterativos. Tienen como ventaja la construcción de software sentando las bases en la reutilización de código (**Larman, 1999**). Entre los patrones GoF utilizados durante la implementación de la herramienta para el análisis estadístico de los datos geológico - mineros se encuentra:

- Singleton: representa la estructura única de una clase a través de una sola instancia de esta, soportando la visibilidad total (mediante un solo punto de acceso) desde cualquier lugar del sistema y, delegando las funciones pertinentes y el cumplimiento de las diversas responsabilidades a ese único objeto (**Larman, 1999**). En la herramienta para el análisis estadístico de los datos geológico - mineros se pone de manifiesto el uso del patrón Singleton en la clase ConectarseBD, debido a que es necesario tener sólo una instancia de esta clase para conservar el estado de la conexión. De esta manera se evita tener varias conexiones innecesarias a la base de datos lo que acarrea ineficiencia y uso abusivo de recursos.

### Diagrama de clases del diseño

En los Diagramas de Clases del Diseño (DCD) es donde se representan las clases del diseño y sus relaciones, atributos y métodos.

---

<sup>9</sup> Pandilla de los Cuatro

Diagrama de clases del diseño para el caso de uso del sistema Calcular estadígrafos.

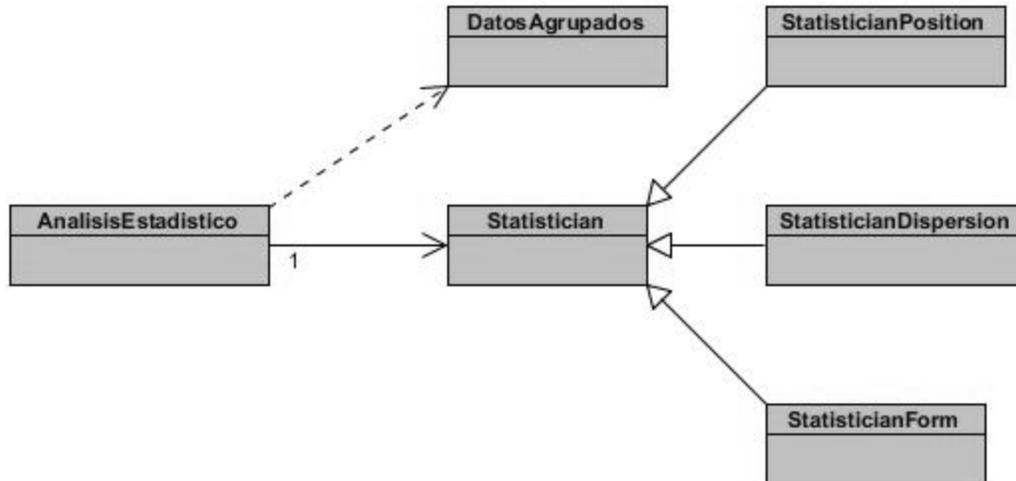


Ilustración 5 Diagrama de clases del diseño del CUS Calcular estadígrafos

Los diagramas de clases del diseño de los CUS Generar tabla de frecuencias, Calcular frecuencias, Generar gráficas, Capturar datos y Generar reportes, están recogidos en el **Anexo 2**.

### Descripción de las clases del diseño

Descripción de la clase del diseño AnalisisEstadistico

Descripción de la clase "AnalisisEstadistico"	
<b>Nombre:</b> AnalisisEstadistico	
<b>Tipo de clase:</b> Controladora	
<b>Descripción:</b> Encargada de recibir los datos de la vista y crear los objetos con la información necesaria para que estos realicen sus responsabilidades.	
Atributos	Tipo
_arDatosColumnas	double[]
_longitudArreglo	int

_modelo	QSqlQueryModel
_posicion	StatisticianPosition
_forma	StatisticianForm
_dispersion	StatisticianDispersion
_contador	int
_listFrecuencia	QList<int>
_listIntervalos	QList<double>
_primerQ	double
_mediana	double
_tercerQ	double
_primerQDA	double
_medianaDA	double
_tercerQDA	double
Responsabilidades	
Nombre	Descripción
datosColumnaSeleccionada(QsqlQueryModel *pModelo): void	Obtiene los datos de la tabla y campo seleccionada.
primerCuartil(): double	Devuelve el primer cuartil de los datos
segundoCuartil(): double	Devuelve el segundo cuartil de los datos
tercerCuartil(): double	Devuelve el tercer cuartil de los datos
mediaAritméticaValores(): double	Devuelve la media aritmética de los datos
medianaValores(): double	Devuelve la mediana de los datos

minimoValores(): double	Devuelve el menor de los datos
maximoValores(): double	Devuelve el mayor de los datos
varianzaValores(): double	Devuelve la varianza de los datos
desviacionEstandar(): double	Devuelve la desviación estándar de los datos
coeficienteCurtosis(): double	Devuelve el coeficiente de Curtosis de los datos
coeficienteAsimetria(): double	Devuelve el coeficiente de asimetría de los datos
desviacionAbsoluta(): double	Devuelve la desviación absoluta de los datos
mostrarDatos(): void	Muestra al analista de datos los resultados de los estadígrafos y la tabla de frecuencias
reportePDF(): void	Genera un reporte (con el último análisis estadístico hecho sobre los datos) en formato “.pdf”
on_actionGraficas_triggered(): void	Muestra los gráficos generados con los datos geológico - mineros para una interpretación visual de características tales como: asimetría, simetría, dispersión, entre otras.

Tabla 4 Descripción de la clase del diseño AnalisisEstadistico.

Las descripciones de las clases del diseño están recogidas en el **Anexo 3**.

## Conclusiones parciales

- No es posible la captura de los requisitos y la construcción del software correcto, sin un firme conocimiento del contexto en que se emplaza el sistema. Es por ello que para expresar dicho contexto, de manera tal que sirva de ayuda a los desarrolladores, se realizó el modelo de dominio, proporcionando una mejor comprensión de las entidades con mayor relevancia.
- Cuando un equipo de desarrollo se enfrenta al proceso de construcción de un sistema, automáticamente se adentra en una situación difícil. Esto es debido a que los desarrolladores no van a ser los futuros usuarios del software, y por lo tanto no saben con exactitud que tan bien funciona y si satisface las expectativas de los clientes. Para resolver esto se realizó la captura de los requisitos, ilustrando de manera general el funcionamiento interno del software, expresado en los casos de usos del sistema.
- Los diagramas de clases del diseño constituyen herramientas muy útiles cuando se inicia el flujo de trabajo **Implementación**, ya que estos diagramas proveen una guía necesaria para la planificación de las actividades de los desarrolladores del equipo de trabajo.

## Capítulo 3: Implementación y pruebas

Una vez obtenido el resultado del diseño se pasa a la implementación del sistema en términos de ficheros de código fuente y binario, scripts, ejecutables, entre otros. El propósito principal de la implementación es desarrollar la mayor parte de la arquitectura del sistema y, el sistema como un todo. Además, el modelo de implementación es la entrada principal de las etapas de prueba en donde, cada construcción (tanto interna como intermedia) generada, es sometida a pruebas de integración y, posiblemente también a pruebas de sistema. Además, en la prueba se verifican las versiones finales del software a ser entregados a terceros (**Jacobson, y otros, 2004**).

En el presente capítulo se realiza la propuesta del modelo de implementación y de pruebas correspondiente a la herramienta para el análisis estadístico básico de los datos geológico – mineros de los pozos de perforación. Se definen el diagrama de despliegue, los diagramas de componentes y, se realizan las pruebas a la herramienta para corroborar que satisface las expectativas.

### Modelo de implementación

El modelo de implementación describe cómo los elementos del modelo del diseño se implementan en términos de componentes. Describe también cómo se organizan los componentes de acuerdo con los mecanismos de estructuración y modularización (habilitados en el ambiente de implementación) y en el lenguaje o lenguajes de programación utilizados, y cómo dependen los componentes uno de otros. Se representa con un sistema de implementación que define el subsistema de nivel superior del modelo. Se puede utilizar otros subsistemas como una forma de organizarlo en partes más manejables (**Jacobson, y otros, 2004**).

### Diagrama de despliegue

El diagrama de despliegue es utilizado como entrada fundamental en las actividades del diseño e implementación debido a que la distribución del sistema tiene una influencia principal en su diseño. Es un modelo de objetos que describe la distribución física de las funcionalidades del sistema entre los nodos de cómputo (**Jacobson, y otros, 2004**).

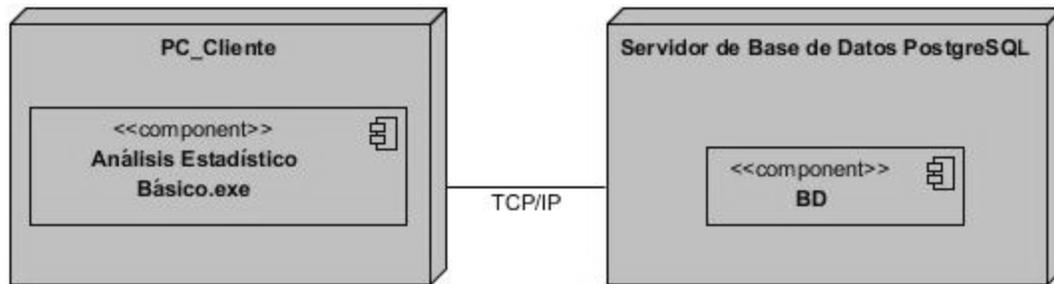


Ilustración 6 Diagrama de Despliegue

### Componente

Un componente es el empaquetamiento físico de los elementos de un modelo, como las clases en el modelo de diseño (Eriksson, y otros, 1998).

### Diagrama de componentes

Los diagramas de componentes describen los elementos físicos y sus relaciones en el entorno. Además, muestran las opciones de realización (Visconti, y otros, 2004).

Diagrama de componentes del CUS Calcular estadígrafos:

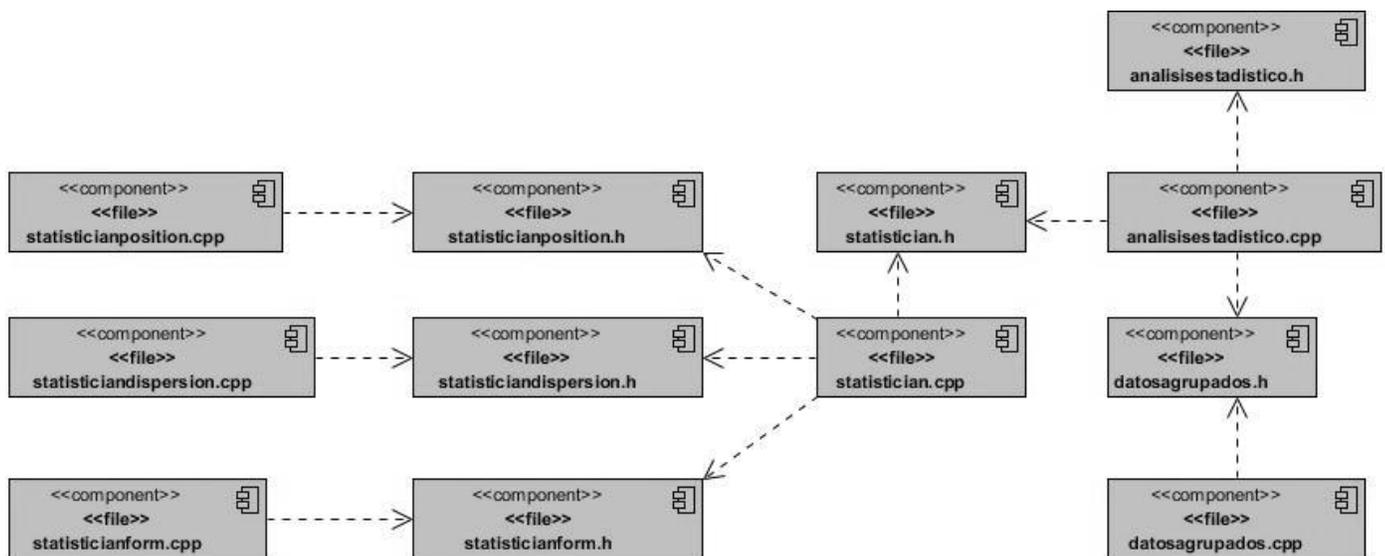


Ilustración 7 Diagrama de componentes del CUS Calcular estadígrafos

Los diagramas de componentes de los CUS Generar tabla de frecuencias, Calcular frecuencias, Generar gráficas, Capturar datos y Generar reportes, están recogidos en el **Anexo 4**.

## Modelo de pruebas

El modelo de pruebas describe principalmente cómo se prueban los componentes ejecutables en el modelo de implementación. También puede describir cómo han de ser probados aspectos específicos del sistema como por ejemplo: si la interfaz de usuario es utilizable y consistente (**Jacobson, y otros, 2004**).

En el presente trabajo se realizarán pruebas de especificación ("caja negra") y de estructura ("caja blanca"), con el objetivo de verificar el comportamiento de la unidad observable externamente y la interacción interna entre los componentes del sistema.

### Caso de prueba

Especifica una forma de verificar el software, incluyendo la entrada o resultado con la que se ha de probar y las condiciones bajo las que ha de probarse. En la práctica, lo que se prueba puede estar dado por uno o varios requisitos funcionales del sistema, cuya implementación justifica una comprobación que es posible realizar y que no es demasiado cara de ejecutar (**Jacobson, y otros, 2004**).

El siguiente caso de prueba corresponde a la prueba de especificación.

Caso de prueba del CUS Calcular estadígrafos:

Entrada
Existe un conjunto de datos perteneciente a la tabla <i>tb_assay</i> y columna <i>gold</i> formado por: <ul style="list-style-type: none"> <li>➤ 4813 muestras que toman valores entre 0.01 y 135.</li> </ul>
Salida
<ul style="list-style-type: none"> <li>➤ El total de datos es: 4813.</li> <li>➤ El valor mínimo de los datos es: 0.01.</li> <li>➤ El valor máximo de los datos es: 135.</li> <li>➤ El rango de los datos es: 134.99.</li> </ul>

- El primer cuartil de los datos sin agrupar es: 0.04, y agrupados es: 2.62664.
- El segundo cuartil de los datos sin agrupar es: 1, y agrupados es: 5.24545.
- El tercer cuartil de los datos sin agrupar es: 1, y agrupados es: 7.86427.
- El rango intercuartil de los datos sin agrupar es: 0.96, y agrupados es: 5.23763.
- La media aritmética de los datos sin agrupar es: 1.15323, y agrupados es: 5.45435.
- La media geométrica de los datos sin agrupar es: 1, y agrupados es: 1.
- La mediana de los datos sin agrupar es: 1, y agrupados es: 5.24545.
- El centro de amplitud de los datos es: 67.495.
- La varianza de los datos sin agrupar es: 18.6621, y agrupados es: 33.2889.
- La desviación estándar de los datos sin agrupar es: 4.31997, y agrupados es: 5.76965.
- El coeficiente de variación de los datos sin agrupar es: 3.74598, y agrupados es: 1.05781.
- El coeficiente de Curtosis de los datos sin agrupar es: 485.454, y agrupados es: 437.212.
- El coeficiente de asimetría de los datos sin agrupar es: 19.6743, y agrupados es: 19.2206.
- La desviación absoluta de los datos sin agrupar es: 1.07196, y agrupados es: 0.50044.

**Condiciones**

Se han obtenido los datos mediante el CUS Capturar datos.

*Tabla 5 Caso de prueba del CUS Calcular estadígrafos*

Los casos de prueba de especificación de los CUS Generar tabla de frecuencias, Generar gráficas, Capturar datos y Generar reportes, están recogidos en el **Anexo 5**.

**Procedimiento de prueba**

Un procedimiento de prueba es una instrucción para un individuo sobre cómo se debe realizar un caso de prueba manualmente, o puede ser una especificación de cómo interaccionar manualmente con una herramienta de automatización de pruebas para crear componentes de prueba (**Jacobson, y otros, 2004**).

El siguiente procedimiento de prueba corresponde a la prueba de especificación.

Procedimiento de prueba del CUS Calcular estadígrafos:

Procedimiento
<ol style="list-style-type: none"> <li>1. Seleccione "Obtener datos" de la ventana principal.</li> <li>2. Verifique que:                             <ul style="list-style-type: none"> <li>➤ El total de datos es: 4813.</li> <li>➤ El valor mínimo de los datos es: 0.01.</li> <li>➤ El valor máximo de los datos es: 135.</li> <li>➤ El rango de los datos es: 134.99.</li> <li>➤ El primer cuartil de los datos sin agrupar es: 0.04, y agrupados es: 2.62664.</li> <li>➤ El segundo cuartil de los datos sin agrupar es: 1, y agrupados es: 5.24545.</li> <li>➤ El tercer cuartil de los datos sin agrupar es: 1, y agrupados es: 7.86427.</li> <li>➤ El rango intercuartil de los datos sin agrupar es: 0.96, y agrupados es: 5.23763.</li> <li>➤ La media aritmética de los datos sin agrupar es: 1.15323, y agrupados es: 5.45435.</li> <li>➤ La media geométrica de los datos sin agrupar es: 1, y agrupados es: 1.</li> <li>➤ La mediana de los datos sin agrupar es: 1, y agrupados es: 5.24545.</li> <li>➤ El centro de amplitud de los datos es: 67.495.</li> <li>➤ La varianza de los datos sin agrupar es: 18.6621, y agrupados es: 33.2889.</li> <li>➤ La desviación estándar de los datos sin agrupar es: 4.31997, y agrupados es: 5.76965.</li> </ul> </li> </ol>

- El coeficiente de variación de los datos sin agrupar es: 3.74598, y agrupados es: 1.05781.
- El coeficiente de Curtosis de los datos sin agrupar es: 485.454, y agrupados es: 437.212.
- El coeficiente de asimetría de los datos sin agrupar es: 19.6743, y agrupados es: 19.2206.
- La desviación absoluta de los datos sin agrupar es: 1.07196, y agrupados es: 0.50044.

*Tabla 6 Procedimiento de prueba del CUS Calcular estadígrafos*

Los procedimientos de prueba de los CUS Generar tabla de frecuencias, Generar gráficas, Capturar datos y Generar reportes, están recogidos en el **Anexo 6**.

La prueba de estructura realiza un estudio cuidadoso de los detalles procedimentales, chequeando los itinerarios lógicos del programa, examinando el estado del software en varios puntos, y comprobando los bucles y condiciones (**Jacobson, y otros, 2004**). Se realizó un estudio de los caminos lógicos más importantes para evitar prolongar los planes de desarrollo de la herramienta, ya que definir todos los caminos lógicos, desarrollar casos de prueba para cada uno de estos y, evaluar los resultados, supone un estudio demasiado exhaustivo.

El método void seleccionDatosAtipicos () es el encargado de encontrar los valores de las muestras que se encuentran numéricamente distante del resto, es decir, los valores que son mucho más pequeños o grandes que el resto de los valores del conjunto de datos. Estos pueden ser indicativos de la existencia de varias poblaciones en la muestra de datos establecida.

El siguiente caso de prueba corresponde a la del camino básico.

Clase Graficas. Función void seleccionDatosAtipicos ().

```

void Graficas::seleccionDatosAtipicos()
{
    double longitudBigote = 1.5 * (_tercerQ - _primerQ);
    double minimo = _primerQ - longitudBigote;
    double maximo = _tercerQ + longitudBigote;

    for(int i = 0; i < _longitud; i++)
    {
        if(_arDatos[i] >= minimo && _arDatos[i] <= _primerQ){
            _minimo = _arDatos[i];
            break;
        }
        else{
            _minimo = _primerQ;
            if(_arDatos[i] < minimo){
                _listAtipicos.append(_arDatos[i]);
            }
        }
    }
    for(int i = _longitud - 1; i > 0; i--)
    {
        if(_arDatos[i] <= maximo && _arDatos[i] >= _tercerQ){
            _maximo = _arDatos[i];
            break;
        }
        else{
            _maximo = _tercerQ;
            if(_arDatos[i] > maximo){
                _listAtipicos.append(_arDatos[i]);
            }
        }
    }
}

```

Ilustración 8 Método seleccionDatosAtipicos

Paso #1: Grafo de flujo de datos.

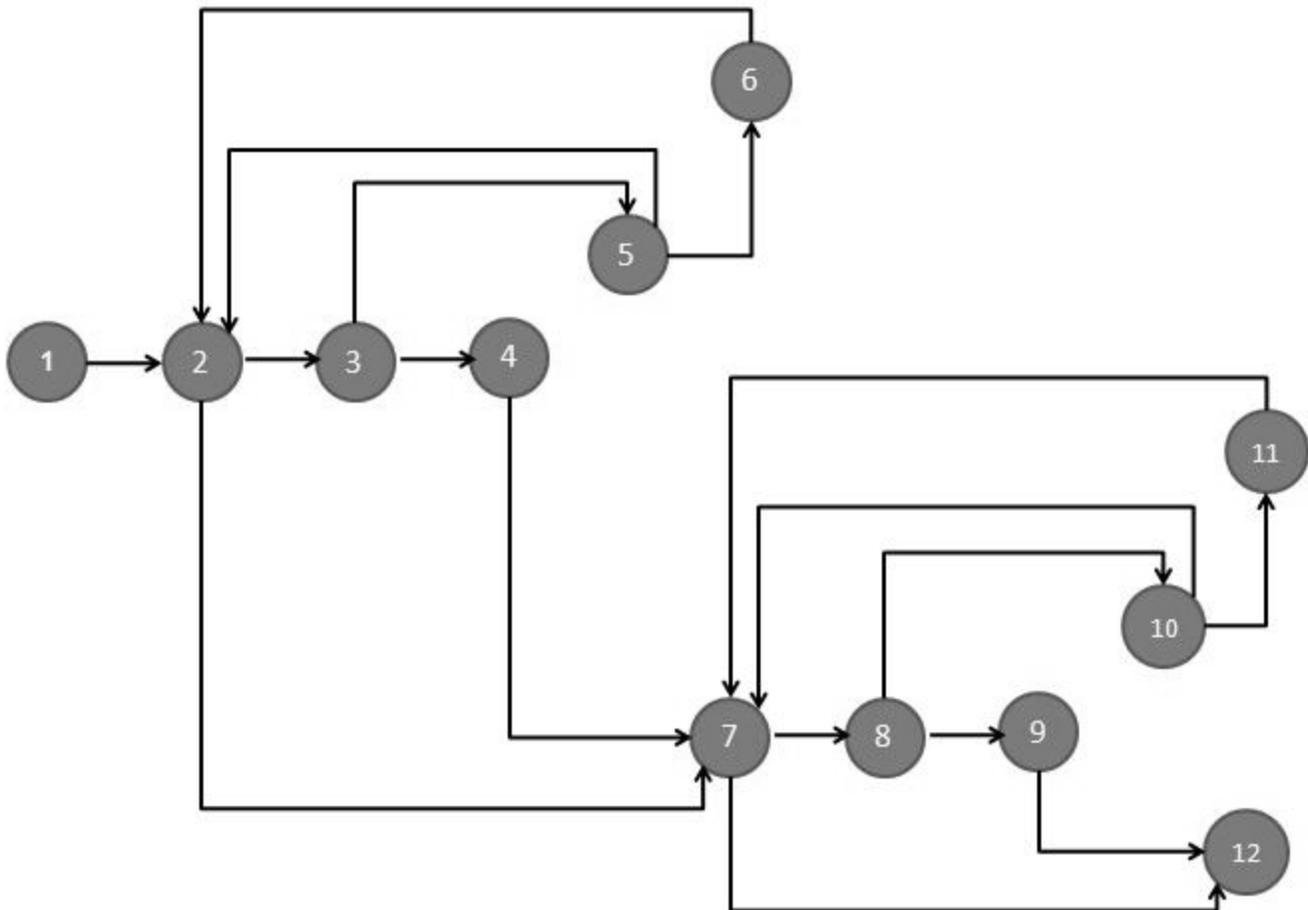


Ilustración 9 Grafo de flujo de datos para el método *seleccionDatosAtipicos*

Paso #2: Complejidad ciclomática.

$$V(G) = 17 - 12 + 2$$

$$V(G) = 7$$

Paso #3: Caminos básicos.

- 1-2-7-12
- 1-2-7-8-9-12

- 1-2-7-8-10-11-7-12
- 1-2-3-4-7-12
- 1-2-3-5-6-2-7-12
- 1-2-7-8-10-7-12
- 1-2-3-5-2-7-12

Paso #4: Casos de pruebas.

Caso de Prueba #1	
<b>Camino:</b>	1-2-7-12
<b>Caso de Prueba:</b>	Probar la función void seleccionDatosAtipicos()
<b>Entrada:</b>	Primer cuartil = 0.04, segundo cuartil = 1, tercer cuartil = 1, valor mínimo = 0.01, valor máximo = 135, conjunto de datos pertenecientes a la base de datos <i>graficar_validaciones</i> , tabla <i>tb_assay</i> y columna <i>gold</i> .
<b>Resultado:</b>	Seleccionar los datos atípicos.

Tabla 7 Caso de prueba para el método seleccionDatosAtipicos

Caso de Prueba #2	
<b>Camino:</b>	1-2-7-8-9-12
<b>Caso de Prueba:</b>	Probar la función void seleccionDatosAtipicos()
<b>Entrada:</b>	Primer cuartil = 0.04, segundo cuartil = 1, tercer cuartil = 1, valor mínimo = 0.01, valor máximo = 135, conjunto de datos pertenecientes a la base de datos <i>graficar_validaciones</i> , tabla <i>tb_assay</i> y columna <i>gold</i> .
<b>Resultado:</b>	Seleccionar los datos atípicos.

Tabla 8 Caso de prueba para el método seleccionDatosAtipicos

Caso de Prueba #3	
<b>Camino:</b>	1-2-7-8-10-11-7-12
<b>Caso de Prueba:</b>	Probar la función void seleccionDatosAtipicos()
<b>Entrada:</b>	Primer cuartil = 0.04, segundo cuartil = 1, tercer cuartil = 1, valor mínimo = 0.01, valor máximo = 135, conjunto de datos pertenecientes a la base de datos <i>graficar_validaciones</i> , tabla <i>tb_assay</i> y columna <i>gold</i> .
<b>Resultado:</b>	Seleccionar los datos atípicos.

Tabla 9 Caso de prueba para el método seleccionDatosAtipicos

Caso de Prueba #4	
<b>Camino:</b>	1-2-3-4-7-12
<b>Caso de Prueba:</b>	Probar la función void seleccionDatosAtipicos()
<b>Entrada:</b>	Primer cuartil = 0.04, segundo cuartil = 1, tercer cuartil = 1, valor mínimo = 0.01, valor máximo = 135, conjunto de datos pertenecientes a la base de datos <i>graficar_validaciones</i> , tabla <i>tb_assay</i> y columna <i>gold</i> .
<b>Resultado:</b>	Seleccionar los datos atípicos.

Tabla 10 Caso de prueba para el método seleccionDatosAtipicos

Caso de Prueba #5	
<b>Camino:</b>	1-2-3-5-6-2-7-12
<b>Caso de Prueba:</b>	Probar la función void seleccionDatosAtipicos()

<b>Entrada:</b>	Primer cuartil = 0.04, segundo cuartil = 1, tercer cuartil = 1, valor mínimo = 0.01, valor máximo = 135, conjunto de datos pertenecientes a la base de datos <i>graficar_validaciones</i> , tabla <i>tb_assay</i> y columna <i>gold</i> .
<b>Resultado:</b>	Seleccionar los datos atípicos.

Tabla 11 Caso de prueba para el método *seleccionDatosAtipicos*

Caso de Prueba #6	
<b>Camino:</b>	1-2-7-8-10-7-12
<b>Caso de Prueba:</b>	Probar la función void <i>seleccionDatosAtipicos()</i>
<b>Entrada:</b>	Primer cuartil = 0.04, segundo cuartil = 1, tercer cuartil = 1, valor mínimo = 0.01, valor máximo = 135, conjunto de datos pertenecientes a la base de datos <i>graficar_validaciones</i> , tabla <i>tb_assay</i> y columna <i>gold</i> .
<b>Resultado:</b>	Seleccionar los datos atípicos.

Tabla 12 Caso de prueba para el método *seleccionDatosAtipicos*

Caso de Prueba #7	
<b>Camino:</b>	1-2-3-5-2-7-12
<b>Caso de Prueba:</b>	Probar la función void <i>seleccionDatosAtipicos()</i>
<b>Entrada:</b>	Primer cuartil = 0.04, segundo cuartil = 1, tercer cuartil = 1, valor mínimo = 0.01, valor máximo = 135, conjunto de datos pertenecientes a la base de datos <i>graficar_validaciones</i> , tabla <i>tb_assay</i> y columna <i>gold</i> .
<b>Resultado:</b>	Seleccionar los datos atípicos.

Tabla 13 Caso de prueba para el método *seleccionDatosAtipicos*

Resultado de la prueba.

Objetivo	Caso de Prueba	Resultado
Encontrar los valores atípicos de la tabla <i>tb_assay</i> y columna <i>gold</i> .	1.	No cumple
	2.	Satisfactorio
	3.	Satisfactorio
	4.	Satisfactorio
	5.	Satisfactorio
	6.	Satisfactorio
	7.	Satisfactorio

Tabla 14 Resultados de los casos de prueba

La prueba del camino básico correspondiente a la función `QList<double> mediaAritmeticalIntervalos()` se encuentra en el **Anexo 7**.

### Resultados de las pruebas

Los casos de prueba fueron diseñados de tal forma que tuvieran una alta probabilidad de mostrar diferentes clases de errores, haciéndolo con la menor cantidad de tiempo y de esfuerzo. Como resultado de la aplicación de las pruebas de especificación y de estructura, se descubrieron errores no detectados hasta entonces, como por ejemplo: cuando se construyó la tabla de distribución de frecuencias y se determinaban la cantidad de muestras por intervalos, en el último intervalo ocurría que si un determinado valor coincidía con el límite superior, este no era considerado como miembro del intervalo. Los errores fueron fáciles de corregir, motivo por el cual la calidad y la fiabilidad de la herramienta para el análisis estadístico básico de los datos geológico – mineros son aceptables

## Conclusiones parciales

- Una vez terminado los trabajos relacionados con el diseño, los cuales permiten capturar la mayor parte de la arquitectura del software, se procede al desarrollo de dicha arquitectura y el sistema como un todo. Esto es posible gracias al papel que juega la implementación en el ciclo de vida del software, el cual se centra en implementar los elementos del diseño, como por ejemplo las clases, en términos de componentes, de forma tal que se encuentren organizados y, se garantice la dependencia entre estos. Además, la descripción del modelo de implementación ilustra de manera efectiva el funcionamiento interno de la aplicación y la implementación de los requisitos funcionales en términos de componentes.
- Para que el sistema a entregar cumpla con los requisitos definidos y se obtengan buenos resultados en cuanto a aceptación mostrado por los clientes; se desarrolle un producto confiable, robusto y cumpla con los requisitos mínimos de calidad, es necesario probar el sistema en diferentes escenarios, sean favorables o desfavorables. Esto se logra verificando el resultado de la implementación de cada construcción. Por tal motivo se realizaron las pruebas de especificación y de estructura, las cuales permitieron encontrar los errores cometidos durante la implementación. Una vez corregidos los errores se demuestra que el software es completamente funcional y confiable.

## Conclusiones generales

Durante el desarrollo del presente trabajo se arribaron a las siguientes conclusiones:

- Con la utilización de la herramienta desarrollada, las tareas de análisis estadístico que conllevaban tiempo y esfuerzo por parte del analista de datos, se lograrán realizar con mayor rapidez y menor esfuerzo. Los cálculos y las representaciones gráficas se tornan más confiables, requisito indispensable para poder tomar decisiones complejas que guíen las operaciones en el sector minero. La solución informática desarrollada, hace que el trabajo del analista de datos se torne más simple, y que sean más reales los resultados que se obtengan a partir de su uso.
- Los artefactos generados en el proceso de desarrollo de la herramienta servirán para que otros desarrolladores obtengan un entendimiento de su estructura, facilitando la realización de posibles modificaciones o la agregación de nuevas funcionalidades al mismo.
- El software fue desarrollado utilizando herramientas multiplataforma y teniendo en cuenta los principios de soberanía tecnológica que defiende el país. En las empresas y compañías geomineras las herramientas con las que se trabajan constituyen software propietarios o de procedencia nacional, las cuales no cubren todas las necesidades a las que se enfrenta la industria minera. Por tal motivo, el uso de herramientas libres propicia como ventaja principal, el mínimo de gastos en cuanto al pago de licencias de software, capacitación y actualización, y además la ampliación de las actividades relacionadas al minado.

## Recomendaciones

Teniendo en cuenta los resultados alcanzados durante la investigación, se proponen las siguientes recomendaciones:

- Realizar el análisis estadístico descriptivo bivariado para analizar las relaciones entre varios atributos de las muestras tomadas de los pozos de perforación.
- Realizar el análisis estadístico descriptivo espacial para identificar características espaciales de los datos geológico – mineros de los pozos de perforación.

## Bibliografía

- Addinsoft. 2013.** XSLTAT. [En línea] 2013. [Citado el: 10 de Enero de 2013.] <http://www.xlstat.com/es/>.
- Alarcón, Raúl. 2000.** *Diseño Orientado a Objetos UML*. Madrid : EIDOS, 2000.
- Alfonso Roche, José R. 1989.** *Estadística en las Ciencias Geológicas*. La Habana : ENSPES, 1989.
- ALGLIB Project. 2013.** ALGLIB. [En línea] 2013. [Citado el: 26 de Marzo de 2013.] <http://www.alglib.net/>.
- Amaya, S. 2013.** cplusplus. [En línea] 3.1, 2013. [Citado el: 10 de Enero de 2013.] Usuario del sitio: Albatross. <http://www.cplusplus.com/info/description/>.
- Billy Reynoso, Carlos. 2004.** *Introducción a la arquitectura de software*. Buenos Aires : Universidad de Buenos Aires, 2004.
- Booch, Grady, Rumbaugh, James y Jacobson, Ivar. 2000.** *Proceso Unificado de Desarrollo de Software*. Madrid : Editorial Addison Wesley , 2000. ISBN: 74-7829-036-2.
- Cassiraga, Eduardo. 2005.** *Curso de geoestadística*. Valencia : Universidad Politécnica de Valencia, 2005.
- Cortizo Pérez, José Carlos, Expósito Gil, Diego y Ruiz Leyva, Miguel. 2004.** *Extreme Programming 1*. España : s.n., 2004.
- Cruz, Elmidio Estévez. 2009.** *Apuntes sobre estimación de recursos y reservas*. 2009.
- de la Torre Llorente, César, y otros. 2010.** *Guía de Arquitectura N-Capas orientada al Dominio con .NET 4.0*. s.l. : Krasis Consulting, 2010. 978-84-936696-3-8.
- Departamento Administrativo Nacional de Estadística. 2011.** *Caracterización Temática Minero Energético*. Bogotá : Dirección de Regulación, Planeación, Estandarización y Normalización, 2011.
- Eriksson, Hans Erik y Penker, Magnus. 1998.** *UML Toolkit*. New York : John Wiley & Sons, 1998.
- Franco Lemus, Miller Giovanni y Gil Sánchez, Cristian Gerardo. 2005.** *Análisis Estadístico en Matlab*. Bogotá : Universidad Nacional Colombia, 2005.
- Free Software Foundation. 2012.** GNU Operating System. [En línea] 31 de Enero de 2012. [Citado el: 1 de Enero de 2013.] <http://www.gnu.org/software/gsl/>.
- . **2013.** Sistema Operativo GNU. [En línea] 2013. [Citado el: 8 de Enero de 2013.] [http://www.gnu.org/software/pssp/manual/html\\_node/Statistical-Functions.html#Statistical-Functions](http://www.gnu.org/software/pssp/manual/html_node/Statistical-Functions.html#Statistical-Functions).
- FUNDIBEQ. 2010.** Fundación Iberoamericana para la Gestión de la Calidad. [En línea] 2010. [Citado el: 4 de 12 de 2012.] <http://www.fundibeq.org/opencms/export/sites/default/PWF/downloads/gallery/methodology/tools/histograma.pdf>.
- Garcilaso, Jordana. 2008.** *Introducción Open Up*. [En línea] Octubre de 2008. [Citado el: 14 de Enero de 2013.] [http://www.mug.org.ar/Descargas/Jornadas/Downloads\\_GetFile.aspx?id=3136](http://www.mug.org.ar/Descargas/Jornadas/Downloads_GetFile.aspx?id=3136).
- Garzón Pérez, María Teresa. 2010.** *Sistemas Gestores de Bases de Datos*. Granada : s.n., 2010. ISSN 1988-6047.
- Gemcom Software International Inc. 2013.** GEMCOM. [En línea] 2013. [Citado el: 9 de Enero de 2013.] <http://es.gemcomsoftware.com/productos/surpac>.
- . **2013.** GEMCOM. [En línea] 2013. [Citado el: 5 de Enero de 2013.] [http://es.gemcomsoftware.com/sites/default/files/products/surpac/es/SurpacModules\\_SPAN.pdf](http://es.gemcomsoftware.com/sites/default/files/products/surpac/es/SurpacModules_SPAN.pdf).

- Guerrero, Rafael Martínez. 2010.** PostgreSQL-es. [En línea] 2010. [Citado el: 12 de Enero de 2013.] [http://www.postgresql.org.es/sobre\\_postgresql#caracteristicas](http://www.postgresql.org.es/sobre_postgresql#caracteristicas).
- Iglesias, Elizabeth Martínez. 2007.** *Adaptación de la metodología RUP a los nuevos módulos del Proyecto Registros y Notarías*. La Habana : Universidad de las Ciencias Informáticas, 2007.
- Instituto Superior Politécnico José Antonio Echeverría. Elementos para el procesamiento estadístico.** La Habana : Instituto Superior Politécnico José Antonio Echeverría.
- Jacobson, Ivar, Booch, Grady y Rumbaugh, James. 2004.** *El Proceso Unificado de Desarrollo de Software*. La Habana : Felix Varela, 2004. Vol. 1. 0-201-57169-2.
- Johnson, Robert y Kubly, Patricia. 2003.** *Estadística elemental: lo esencial*. Tercera edición. 2003.
- Larman, Craig. 1999.** *UML y Patrones Introducción al análisis y diseño orientado a objetos*. México : Prentice Hall, 1999. 970-17-0261-1.
- Martell, Vladimir. 2009.** *Proyecto Técnico Sistema de Datos Geológicos*. La Habana : Geoinformática, 2009.
- Martín Pilego, Francisco Javier y García Secades, Marta. 2007.** *Introducción a la Estadística Económica y Empresarial*. Tercera edición. Madrid : Thomson, 2007.
- Medina Patón, Eduardo Fernández. 2006.** Alarcos. [En línea] 2006. [Citado el: 1 de Febrero de 2013.] <http://alarcos.inf-cr.uclm.es/doc/ISOFTWAREI/Tema04.pdf>.
- Meyer, Lisandro Damián Nicanor Pérez. 2007.** [En línea] 2007. [Citado el: 22 de Enero de 2013.] <http://perezmeyer.com.ar/files/introduccionAQ/Introducci%3n%20a%20Qt4.pdf>.
- Minitab Inc. 2013.** Software for Quality Improvement. [En línea] 2013. [Citado el: 10 de Enero de 2013.] <http://www.minitab.com/es-MX/products/minitab/>.
- Ortiz Noriega, Dresky y López Boullón, Carlos Javier. 2010.** *Módulo de REKO para la resolución de conflictos*. La Habana : Universidad de las Ciencias informáticas, 2010.
- Pavón Mestras, Juan. 2008.** *El patrón Modelo-Vista-Controlador*. Madrid : Universidad Complutense Madrid, 2008.
- Pérez López, César. 2001.** *Técnicas estadísticas SPSS*. Primera. s.l. : Pearson, 2001. 48-205-3167-7.
- Pérez, María Teresa Garzón. 2010.** *Sistemas Gestores de Bases de Datos*. Granada : s.n., 2010. 1988-6047.
- Polanco Almanza, Ramón Gilberto, Batista Leyva, Joel y Legra Lobaina, Arístides Alejandro. 2001.** *Informatización de la minería en la industria cubana del níquel*. Moa : Instituto Superior Minero Metalúrgico, 2001. 0258 5979.
- Pressman, Roger S. 2002.** *Ingeniería del Software: un enfoque práctico*. Quinta. s.l. : McGraw-Hill Companies, 2002. 84-481-3214-9.
- Pressman, Roger S. 2000.** *El Proceso Unificado de Desarrollo de Software*. Madrid : Pearson Educación S.A., 2000. 84-7829-036-2.
- . **2007.** *Ingeniería de Software: un enfoque práctico*. Sexta edición. Nueva York : McGraw-Hill, 2007. pág. 67.
- Qt Project Hosting. 2011.** Qt Project. [En línea] 2011. [Citado el: 12 de Enero de 2013.] <http://qt-project.org/wiki/Category:Tools::QtCreator>.
- Ruiz, Héctor José Rodríguez. 2008.** *Historia sobre el uso de las técnicas informáticas en la actividad geológica - minera en Cuba*. Santiago de Cuba : Empresa Geominera de Oriente, 2008.

- Salazar, Norberto Guarín. 2002.** *Estadística Aplicada*. Medellín : Universidad Nacional de Colombia, 2002.
- Salvador Figueras, Manuel y Gargallo Valero, Pilar. 2003.** *Análisis Exploratorio de Datos*. Zaragoza : Universidad de Zaragoza, 2003.
- Sironvalle, Marco Antonio Alfaro. 2000.** *Estadística*. 2000.
- . **2002.** *Introducción al Muestreo Minero*. Santiago de Chile : Instituto de Ingenieros de Minas de Chile, 2002.
- SolMine. 2013.** Soluciones Mineras Libres SRL. *SolMine*. [En línea] 2013. [Citado el: 31 de Marzo de 2013.]
- Sommerville, Ian. 2005.** *Ingeniería del Software*. Madrid : Pearson Educación S.A., 2005.
- The R Project for Statistical Computing. 1997.** R Foundation. [En línea] 1997. [Citado el: 1 de Enero de 2013.] <http://www.r-project.org/>.
- Villar, Malay Rodríguez. 2007.** *Introducción de procedimientos ágiles en la producción de software en la Facultad 7 de la Universidad de las Ciencias Informáticas*. La Habana : Universidad de las Ciencias Informáticas, 2007.
- Visconti, Marcello y Astudillo, Hernán. 2004.** Departamento de Informática . [En línea] 22 de Octubre de 2004. [Citado el: 21 de 3 de 2013.] <http://www.inf.utfsm.cl/~visconti/ili236/Documentos/15-Implementacion.pdf>.