UNIVERSIDAD DE CIENFUEGOS ”CARLOS RAFAEL RODRÍGUEZ”
FACULTAD DE INGENIERÍA INFORMÁTICA

SISTEMA INFORMÁTICO PARA EL CÁLCULO DEL TAMAÑO DE MUESTRAS

Tesis de Grado

Autor: MORGAN REYES ÁLVAREZ
Tutores: Prof. Tit., MANUEL CORTÉS CORTÉS, Dr.
Prof. Aux., ELIA CABRERA ÁLVAREZ, MsC.
Prof. Ing., YAILEM ARENCIBIA DEL REY.

Cienfuegos
2008

A mi familia

y amigos

Gracias a todos

y cada uno de ustedes

RESUMEN

El presente trabajo lleva por título: ”SICTM. Sistema informático para el cálculo del tamaño de muestras”, el centro de esta investigación lo constituyen los modelos matemáticos de la teoría del muestreo, vistos como alternativa de solución para la obtención de tamaños de muestra que garanticen la representatividad de los estadísticos expuestos a partir de una población objeto de estudio.

El proceso relacionado con la determinación del tamaños de muestras se desarrolla en su totalidad mediante la confección manual de diferentes reportes asociados a los estadísticos descriptivos, las tablas de números aleatorios que conforman el marco muestral de la investigación. Todo lo anterior se trasluce en una considerable pérdida tiempo y recursos, además de resultar una tarea tediosa y de muy complicado proceder para los experimentadores involucrados en la investigación muestral. Luego, con la introducción de las tecnologías informáticas se facilitaría notablemente todo este proceso.

SICTM, posibilita la obtención de tamaños de muestras de forma rápida garantizando altos niveles de confiabilidad en los resultados, lográndose por consiguiente una reducción de los costos asociados a la investigación. Permite la confección adecuada de reportes en relación a los elementos que conforman la muestra extraída de la población, agilizando el trabajo con las tablas de números aleatorios. Incorpora, a diferencia de los software existentes la posibilidad de realizar un estudio piloto, donde le ofrece al experimentador el comportamiento de los estadísticos poblacionales de forma preliminar y por tanto una medida de viabilidad económica respecto a la investigación en curso. Posee además un sistema de ficheros que almacena de manera eficiente la información requerida, así como, un asistente de ayuda para el uso y comprensión de los conceptos relacionados al sistema informático.

ÍNDICE

Pág.

RESUMEN **III**

ÍNDICE **IV**

INTRODUCCIÓN **1**

1. Fundamentación Teórica **7**

1.1. Introducción 7

1.2. Descripción del dominio del problema 7

1.2.1. Planeación de una encuesta por muestreo 7

1.2.2. Representatividad de la muestra 11

1.3. Diseños fundamentales de la teoría del muestreo 13

1.3.1. Muestreo Irrestrico Aleatorio 13

1.3.2. Muestreo Sistemático 16

1.3.3. Muestreo Aleatorio Estratificado 18

1.3.4. Muestreo Por Conglomerado Monoetápico 23

1.3.5. Muestreo Piloto 26

1.4. Descripción del objeto de estudio 27

1.4.1. Flujo actual de los procesos y análisis crítico de la ejecución de estos . . . 27

1.5. Descripción de los sistemas existentes 28

1.5.1. Algunos software estadísticos a nivel internacional 28

1.5.2. Algunos software estadísticos existentes en Cuba 30

1.5.3. Presentación de la propuesta de solución 30

1.6. Tendencias, metodologías y/o tecnologías actuales 31

1.6.1. Fundamentación de la metodología utilizada 31
1.6.2. Fundamentación del lenguaje y software utilizado 32
1.7. Conclusiones Parciales 34

2. Descripción y construcción de la solución propuesta 35
2.1. Introducción 35
2.2. Descripción del modelo del dominio 36
2.2.1. Definición de entidades y conceptos principales 36
2.2.2. Representación del modelo del dominio 38
2.2.3. Reglas del negocio a considerar 39
2.3. Descripción del sistema propuesto 39
2.3.1. Concepción general del sistema 39
2.3.2. Requerimientos Funcionales 42
2.3.3. Requerimientos No Funcionales 44
2.3.4. Modelo de casos de uso del sistema 46
2.3.5. Diagrama de casos de uso del sistema 48
2.3.6. Descripción de los casos de uso del sistema 49
2.4. Construcción de la solución propuesta 55
2.4.1. Diagrama de clases del diseño 55
2.4.2. Diagrama de archivo 56
2.4.3. Estructura de los archivos 57
2.4.4. Diagrama de implementación 60
2.5. Principios de diseño 60
2.5.1. Estándares en la interfaz de la aplicación 60
2.5.2. Tratamiento de errores 61
2.5.3. Concepción general de la ayuda 61
2.6. Conclusiones Parciales 62

3. Análisis de factibilidad y validación de la solución propuesta 63
3.1. Introducción 63
3.2. Planificación por puntos de función 64
3.3. Determinación de los costos 64

3.4. Beneficios tangibles e intangibles 70
3.5. Análisis de costos y beneficios 71
3.6. Validación de la propuesta de solución 72
3.6.1. Procesamiento estadístico y análisis de los resultados 72
3.6.2. Valoraciones de acuerdo al criterio de expertos 72
3.7. Conclusiones Parciales 79

CONCLUSIONES **80**

RECOMENDACIONES **81**

BIBLIOGRAFÍA **82**

ANEXOS **83**

INTRODUCCIÓN

La dinámica del mundo actual está caracterizada en gran medida por la influencia de la Estadística Matemática dado por su aplicación en diferentes ramas del conocimiento científico tal es el caso de las ciencias sociales, la química, la física, entre otras.

La estadística trata de la selección, análisis y uso de los datos con la finalidad de resolver problemas, es la ciencia de extraer conclusiones acerca de una población con base en un subconjunto de datos tomados de dicha población. Según (Weimer, 1998) es la ciencia de recolectar, organizar, analizar e integrar información. Entiéndase por estadística y probabilidad los métodos para describir la variabilidad, además de permitir la toma de decisiones cuando la variabilidad está presente (Hines, 1998).

Para referirse al término estadística teniendo como punto de partida su extenso campo de aplicación es necesario mencionar que el objeto de estudio de la misma se centra en los *fenómenos aleatorios*. Según el criterio de (Méndez, 1975) todo fenómeno aleatorio la metodología de estudio para éstos es la estadística, que basándose en ciertos postulados probabilísticos, permite hacer descripciones, optimizaciones y predicciones de éstos , es decir fenómenos cuyo resultado final no se puede determinar con suficiente exactitud de manera que se le asocia cierta probabilidad de ocurrencia, además se deben diferenciar claramente sus dos enfoques: la *estadística descriptiva* y la *estadística inferencial*.

En relación a la *estadística descriptiva* es considerada por (Weimer, 1998) la rama de la estadística que trata con la organización, el resumen y la representación de datos, presenta la información

tabulada y en forma de gráficas. En este enfoque están comprendidos aquellos métodos usados para organizar y describir la información recabada, la organización, exploración y resumen de la información contenida en un conjunto de datos grande o pequeño en cantidad.

En la obra de (Mendenhall, 1987) se aborda la *estadística inferencial* como el desarrollo de técnicas para la extracción de inferencias con un grado de error medido en términos probabilísticos. Esta parte de la estadística se encarga también de todos los aspectos relacionados con el muestreo, ya que por definición una inferencia es una generalización acerca de la población con base en los datos de una muestra extraída de ésta (Cochran, 1976). La inferencia estadística construye una medida probabilística de cercanía entre el modelo y los datos. Comprende aquellos métodos y técnicas usados para hacer generalizaciones, predicciones o estimaciones sobre poblaciones a partir de una muestra.

La *teoría del muestreo* estudia la relación entre una población y las muestras tomadas de ella, es útil para determinar si las diferencias entre dos muestras de la misma población se deben a causas fortuitas o son realmente significativas. El análisis de los métodos de muestreo se denomina diseño del experimento y consiste en determinar el tamaño de la muestra, los métodos para seleccionar a los individuos, etc. es decir, el plan de muestreo (Bonilla, 1996).

El muestreo es una herramienta de la investigación científica. Su función básica es determinar que parte de una realidad en estudio (población o universo) debe examinarse con la finalidad de hacer inferencias sobre dicha población. El error que se comete debido al hecho de que se obtienen conclusiones sobre cierta realidad partiéndose de la observación de sólo una parte de ella, se denomina error de muestreo. Obtener una muestra adecuada significa lograr una versión simplificada de la población, que reproduzca de algún modo sus rasgos básicos.

El primer trabajo científico sobre el muestreo de poblaciones finitas se le atribuye al polaco Jerzy Neyman (1934), el cual estableció, sin lugar dudas, que la selección aleatoria es la base de una teoría científica que permite predecir la validez de las estimaciones muestrales, más tarde introduciría la teoría de los intervalos de confianza como una forma de estimar parámetro que determina un intervalo que lo contiene y un coeficiente de confianza que representa la probabilidad

que dicho intervalo efectivamente contenga el parámetro.

El desarrollo y el nivel de aplicación que la Estadística, como herramienta útil y rigurosa en el campo de la investigación ha experimentado en los últimos años, ha sido espectacular. Es indudable que este progreso en el conocimiento y la aplicación de la Estadística ha venido estrechamente vinculado al que ha experimentado el área de la computación, ello como resultado ha llevado a una sociedad absolutamente informatizada donde las TIC (Tecnologías de la Información y la Comunicaciones) se han convertido en un utensilio personal de uso habitual.
El auge y progreso de la informática, a nivel de software y hardware ha hecho posible, a su vez, la realización de pruebas estadísticas que, de forma habitual, hubiesen sido muy costosas desde el punto de vista humano, así como, el manejo de grandes volúmenes de información.

Cuba no está ajena a los avances en las ciencias de la computación, por consiguiente la aplicación de las técnicas de muestreo mediante la perspectiva de utilización de las tecnologías informáticas es sin duda un factor decisivo para el mejoramiento del proceso del diseño de la encuesta. Teniéndose en cuenta esta idea se ha constatado una deficiencia significativa relacionada con la ausencia de un software que gestione la determinación del número de elementos de la muestra, ello se traduce en una pérdida considerable e innecesaria de tiempo y recursos. Actualmente, todo el proceso concerniente al diseño de la encuesta centra su puesta en práctica en el desarrollo manual de los cálculos así como el manejo de las tablas de números aleatorios, por ende, esta labor se torna sumamente tediosa y requiere de gran cuidado por parte de los experimentadores involucrados. De manera adicional se presentan notables dificultades ante la necesidad de aplicar un estudio preliminar sobre la población para corroborar la calidad del modelo de encuesta y determinar parámetros poblacionales que posibiliten elevar la representatividad del tamaño de la muestra y de forma análoga garanticen la confiabilidad de los resultados obtenidos.

Partiéndose de las razones expuestas con anterioridad se identifica como ***problema a resolver*** la necesidad de contar con un *software propio* para el cálculo del tamaño de muestra, utilizando

modelos matemáticos de la teoría del muestreo.

Se considera como ***objeto de estudio*** de la presente investigación los modelos matemáticos de la teoría del muestreo mediante la utilización de las tecnologías informáticas, de este modo se deriva como ***campo de acción*** la determinación efectiva del número de elementos de la muestra.

Se plantea como ***idea a defender*** que el desarrollo de un software para el cálculo del tamaño de muestra posibilita elevar el nivel científico en el área de la investigación y garantizar convenientes costos de inversión y operación durante el proceso de desarrollo de la encuesta por muestreo.

Como ***objetivo general*** se plantea el desarrollo de un software que permita el cálculo del tamaño de muestra, según los diseños fundamentales de la teoría del muestreo.

Para lograr el cumplimiento del *objetivo general* se concertaron los siguientes ***objetivos específicos***:

- Analizar los modelos matemáticos de la teoría del muestreo.
- Diseñar el sistema informático.
- Implementar los diseños fundamentales del muestreo para el cálculo del tamaño de muestra.
- Implementar la determinación de los parámetros poblacionales.
- Confeccionar la ayuda del sistema.

Con la finalidad de alcanzar el cumplimiento de los *objetivos específicos*, se desarrollaron las siguientes ***tareas de investigación***:

- Realizar entrevistas a expertos para conocer el sistema de trabajo, identificar deficiencias o anomalías existentes y valorar las necesidades actuales.
- Recopilar y consultar los documentos relacionados con el tema.
- Seleccionar las herramientas y lenguajes de programación que resulten más efectivos para desarrollar el sistema.
- Identificar los requerimientos del sistema.
- Diseñar e implementar la interfaz gráfica del sistema.

- Preparar el sistema informático.

El presente trabajo de diploma está estructurado en tres capítulos, conclusiones y recomendaciones. El contenido correspondiente a cada capítulo es el siguiente:

Capítulo 1. Fundamentación Teórica:

En esta sección se realiza la descripción del dominio del problema en cuestión, para ello se hace referencia a los aspectos más importantes previstos en una encuesta por muestreo. Se realiza un análisis de la representatividad del tamaño muestral, abordando algunos conceptos de interés. Seguidamente, son introducidos los diseños fundamentales para el cálculo del tamaño de muestra, así como, las estimaciones de los estadísticos de la población por cada técnica muestral. Se han determinado las dificultades y limitantes existentes en la actualidad respecto a la temática en estudio, de igual forma se establecen las tecnologías y metodologías a emplear en el desarrollo del sistema.

Capítulo 2. Descripción y construcción de la solución propuesta:

En este capítulo, se presentan los conceptos fundamentales en la comprensión del entorno del sistema propuesto, se muestra el diagrama de clases del modelo de objetos del dominio. Se establecen las reglas del negocio a considerar , así mismo, quedan determinados los requerimientos funcionales y no funcionales en relación al sistema. Se establece la metodología RUP como herramienta de modelación, se definen los actores y casos de usos involucrados en el sistema, conjuntamente se concreta el diagrama de casos de uso del sistema junto a la correspondiente descripción de cada uno de los mismos. Se presentan el diagrama de clases del diseño, la estructura de los archivos de datos y el diagrama de implementación, además se tienen en cuenta los principios de diseño para la presentación de la interfaz del sistema propuesto, así como, el tratamiento de errores y la concepción de la ayuda del sistema.

Capítulo 3. Análisis de factibilidad y validación de la solución propuesta:

En este capítulo, se describe el proceso de planificación del estudio de factibilidad, se lleva cabo la determinación de los costos, así como, los beneficios tangibles e intangibles asociados al proyecto. Además, se presenta un análisis de los costos y beneficios vinculados al desarrollo de la aplicación.

En relación al procedimiento de validación del sistema, se exponen los criterios de selección de los expertos y se presentan los resultados obtenidos a partir de la aplicación de los modelos de encuesta.

Capítulo 1

Fundamentación Teórica

1.1. Introducción

En el presente capítulo se abordan todos los aspectos relacionados a la fundamentación teórica del tema de investigación, a partir de la descripción los elementos de mayor relevancia en el proceso de planeación de la encuesta. Posteriormente, se analizan los diferentes elementos que inciden en la representatividad del tamaño de la muestra. Se exponen los conceptos relacionados con los diseños fundamentales del muestreo, así como las herramientas matemáticas existentes para cada uno de los mismos. Se abordan las dificultades que se presentan en la actualidad en relación al tema, para lo cual se describen detalladamente los conceptos involucrados en el dominio del problema. Se presentan las soluciones existentes fuera y dentro del ámbito nacional, con el objetivo de establecer la nueva propuesta a partir de los aspectos novedosos respecto a los sistemas existentes. Se realiza, además, una comparación de las tendencias, tecnologías y metodologías actuales, determinándose aquellas que serán utilizadas en el desarrollo del sistema.

1.2. Descripción del dominio del problema

1.2.1. Planeación de una encuesta por muestreo

En esta sección se verán de una manera muy general los aspectos más importantes que intervienen en una encuesta muestral.

- *ESTABLECIMIENTO DE LOS OBJETIVOS*

  Establecer los objetivos de la encuesta de una manera clara y concisa, remitiéndose a estos

conforme se vaya progresando en el diseño e instrumentación de la encuesta. Mantener los objetivos suficientemente simples, de tal manera que sean entendibles por quienes trabajan en la encuesta y sean logrados exitosamente cuando se complete la misma. Es necesario establecer claramente los objetivos de la encuesta pues en los preparatorios de los detalles de ésta es posible que ellos se pierdan de vista y se realicen operaciones encaminadas a obtener objetivos diferentes.

- *POBLACIÓN OBJETIVO*

  La población de la cual se hacen las inferencias, llamada población objetivo, debe definirse lo mas claramente posible. En el transcurso del desarrollo de un estudio, debe tenerse cuidado en asegurar que la población muestreada no se desvíe drásticamente de la población objetivo. Cuando la población muestreada se restringe por conveniencia práctica, se debe tener sumo cuidado al extender las inferencias a la población objetivo.

- *EL MARCO MUESTRAL*

  Se selecciona un marco o varios marcos de tal forma que el listado que la conforma en relación a la población bajo estudio concuerden lo mayor posible. Es importante considerar que el establecimiento de un número considerables de marcos pueden encaminar a una mayor eficiencia.

- *DISEÑO DE MUESTREO*

  Determinar el método de muestreo a ser usado y la selección del tamaño de la muestra son elementos importantes en la planeación de un estudio. La selección de un método de muestreo apropiado se basa en factores como la estructura de la población, el tipo de información buscada, las facilidades administrativas y el personal disponible para ejecutar el estudio. Junto con la elección del método apropiado de muestreo, debe determinarse el tamaño de la muestra especificando el grado de precisión deseado en los estimadores. También debe verificarse si el tamaño elegido de la muestra es factible con el presupuesto destinado para el estudio.

- *MÉTODO DE MEDICIÓN*

  El decidir un método de medición en ocasiones puede estar condicionado a los recursos de que se dispongan; un ejemplo pudiera ser la realización de un estudio de cierta especie de animal en extinción lo cual pudiese requerir de asistencia por parte de un equipo técnico

especializado. Usualmente se emplean algunos de los siguientes métodos de medición: entrevistas por teléfono, entrevistas personales, aplicación de cuestionarios vía telefónica, o a través de correo electrónico, así como la observación directa.

- *INVESTIGADORES DE CAMPO. SELECCIÓN Y ADIESTRAMIENTO*

  Se debe tener gran cuidado a la hora de seleccionar y proceder al adiestramiento de los investigadores de campo. El proceso de instrucción de los investigadores denota una importancia aún mayor si el método de entrevista es personal o por teléfono dado que la presencia de no respuesta, así como la exactitud se verán influenciadas por el estilo del entrevistador.

- *INSTRUCCIONES*

  La preparación de las instrucciones para las diferentes tareas que intervienen en una investigación de muestro es un trabajo difícil, puesto que se necesita un conocimiento profundo del tema de la encuesta, de la psicología del enumerador y del entrevistado, etc.; por ello, es necesario la ayuda de otros especialistas (psicólogos, sociólogos) en esta actividad. Las instrucciones permiten la uniformidad de los procedimientos que genera la encuesta, tales como la obtención de la información por medio de cuestionarios, codificadores, etc. Esta uniformidad minimiza los sesgos que intervienen en la encuesta y que no se deben al modelo matemático.

- *EL CUESTIONARIO*

  La confección del cuestionario es visto muchas veces como una tarea fácil y se piensa erróneamente que su aplicación es independiente al tiempo de preparación de la misma, esto no es así, puesto que se está ante una tarea bastante compleja que requiere mucho estudio, donde inclusive se necesita seleccionar una muestra preliminar con el fin corroborar posibles errores. Debe señalarse que el éxito de una encuesta depende en gran medida de la calidad que tenga el cuestionario.

- *EL ENTREVISTADO*

  Cuando las preguntas que se formulen al entrevistado afecten directa o indirectamente su integridad, autoridad, y dignidad, se debe esperar que sus respuestas estén condicionadas a mantener su prestigio, es decir, puede resultar que dichas propuestas no se ajusten a la realidad. Si las preguntas del cuestionario no están bien redactadas puede suceder que el

entrevistado interprete incorrectamente los objetivos de la encuesta y al información que brinde traiga como consecuencia que los resultados sean sesgados.

- *EL ENUMERADOR*

  Los enumeradores deben ser adiestrados de forma tal que todos tengan la misma interpretación de las preguntas del cuestionario y de los conceptos que intervienen en la encuesta. Si esto se logra, el enumerador puede ayudar eficientemente al entrevistado, aclarándole cualquier mala interpretación a causa de la no compresión correcta de las preguntas del cuestionario. La presencia del enumerador en la casa del entrevistado hace que este se sienta comprometido a participar en la encuesta, aunque no le interesen directamente los objetivos de ella. Si en una encuesta se deja que cada enumerador interprete las preguntas, cada uno hará la suya y todas serán diferentes, lo cual trae como consecuencia que posiblemente los resultados estén seriamente sesgados.

- *ORGANIZACIÓN DEL TRABAJO DE CAMPO*

  Plantearse en detalle el trabajo de campo es de vital importancia para el posterior desempeño de la encuesta ya sea a pequeña o gran escala. En este proceso estará inmerso un número considerable de personal tales como: entrevistadores, coordinadores, o aquel personal destinado al análisis los datos. La definición de los trabajos y el establecimiento de las líneas de dirección antes de dar comienzo a la encuesta debe realizarse de forma consciente, esto propiciará un empleo racionalizado de los recursos dispuestos para el proceso de la encuesta.

- *ORGANIZACIÓN DEL MANEJO DE LOS DATOS*

  Dadas las diferentes etapas que comprende la aplicación de una encuesta, se debe confeccionar un esquema en relación al manejo de la información recolectada. Este plan deberá tener en cuenta los diferentes pasos que comprende el procesamiento de los datos, desde que son medidos en el campo hasta finalizar el análisis de los mismos. Otro factor importante se considera el control de la calidad con la finalidad de verificar la correlación entre los datos procesados y los recolectados en el campo.

- *ANÁLISIS DE DATOS*

  Una vez que se ha completado la encuesta, la fuerza de las técnicas gráficas y numéricas puede ser empleada en su totalidad para interpretar los resultados. Histogramas y tablas de frecuencia conjuntas ayudan a mostrar la correlación entre las respuestas. Inventiva en crear

gráficos puede sugerir interesantes relaciones y conclusiones. En el reporte de los estimativos de las cantidades poblacionales, debe darse la información sobre la incertidumbre en términos de errores probabilísticos o intervalos de confianza.

- *PRECISIÓN Y CONFIABILIDAD*

  Como los resultados de la encuesta provienen de una muestra es posible que se genere cierta incertidumbre alrededor de estos, por eso, es necesario ayudar al usuario a plantear numéricamente cuál es la precisión y confiabilidad que quiere en las estimaciones y a que comprenda estos conceptos.

- *TABLAS*

  Es importante construir el juego de tablas de resultados y que el usuario vea si esa es la información que necesita, para que sirva de base en la confección del cuestionario, a la construcción del modelo matemático y a la tabulación de los datos de la encuesta.

- *ENCUESTA PILOTO*

  Con frecuencia, resulta conveniente emplear una porción del presupuesto en realizar un estudio a pequeña escala llamado prueba piloto. Una prueba piloto es la oportunidad de poner a prueba el cuestionario para detectar y corregir cualquier irregularidad seria o situación inadecuada. Una encuesta piloto también puede dar información o sugerencias para mejorar el diseño del estudio, el mismo constituye un punto de partida capaz de evaluar el cálculo del tamaño muestral. Los resultados aportados por el estudio piloto generalmente sugieren ciertas modificaciones antes de realizar el muestreo a escala completa.

1.2.2. Representatividad de la muestra

La muestra debe reproducir la características del universo, por lo tanto surgen entonces dos preguntas, sobre la cantidad de elementos que debe incluir la muestra y hasta que punto pueden generalizarse a la población. Ambas preguntas convergen en un problema de exactitud o precisión cuya finalidad es no incurrir en errores a la hora de obtener los resultados, no obstante en la práctica los errores son inevitables, luego, lo importante entonces es minimizarlos. Existen dos tipos de errores:

- Los ***sistemáticos o distorsiones***, que son causados por factores externos a la muestra y que se pueden producir en cualquier momento de la investigación.

- El ***error de muestreo***, de ***azar o de estimación***, inevitable, ya que siempre habrá diferencia entre los valores medios de la muestra y los valores medios del universo, la magnitud de este error depende del tamaño de la muestra (a mayor tamaño de muestra menor error) y de la dispersión o desviación (a mayor dispersión mayor error).

Se concluye entonces que para que una muestra sea representativa debe estar dentro de ciertos límites y proporciones establecidas por la estadística. El cálculo del tamaño de la muestra depende de los siguientes elementos:

- Amplitud del universo.

  Se considera finito cuando esta constituido por 100.000 elementos o menos e infinito si excede esta cifra. Cuando se dice que un Universo es infinito entonces la amplitud no influye en el tamaño de la muestra.

- Nivel de confianza adoptado.

  Esta conformado por un límite o intervalo donde figura la probabilidad de que la media de la muestra corresponda a la media de la población. Su cálculo es teórico y se basa en aplicaciones de la ley normal de probabilidad estadística a la distribución en una población de las medias de todas las muestras posibles de un mismo tamaño. En una distribución de este tipo se pueden acotar muchos intervalos, pero para efectos prácticos normalmente se hablan de 3.

  - Entre -1 sigma y +1 sigma s 68.3
  - Entre -2 sigma y +2 sigma s 95.5
  - Entre -3 sigma y +3 sigma s 99.7

- Error de estimación.

  Los resultados obtenidos de la muestra no son rigurosamente exactos con respecto al universo por lo que siempre existirá un margen de error mayor o menor. El error de estimación es siempre inversamente proporcional al tamaño de la muestra, a mayor tamaño menor error. El máximo margen de error que se suele permitir es del 6 por ciento.

- Desviación típica, o de la proporción que se encuentran en el universo las características estudiadas. La proporción en que se encuentra en el universo la característica estudiada. El

tamaño de la muestra depende, de la proporción en porcentaje aproximado o estimado que tiene la población de la característica a estudiar. Se suele expresar p como el porcentaje que posee la característica y q como el porcentaje que no la posee o sea $p + q = 100$.

- Ausencia de distorsión.

  Cuando una muestra presenta alguna anomalía aparecerá sesgada o viciada con errores sistemáticos o de distorsión, lo cual debe evitarse a toda costa. Algunos ejemplos de distorsiones son:

  - Elección deliberada de una muestra no representativa (una encuesta por teléfono, si no la tienen todos los habitantes).
  - Tendencias subjetivas, conscientes o inconsciente del investigador.
  - Sustituciones (encuestas o entrevistas contestadas por el mismo entrevistador).
  - Elección de la muestra dependiente de una característica de las unidades.

1.3. Diseños fundamentales de la teoría del muestreo

1.3.1. Muestreo Irrestrico Aleatorio

La denominación empleada con mayor frecuencia en la literatura es *muestreo aleatorio simple*. No obstante, se encuentran otros sinónimos como *muestreo irrestricto aleatorio* o *muestreo probabilístico de elementos*. Es considerado por (Silva, 1993) como el modelo más simple de la serie y , a su vez la base teórica sobre la que se asienta el resto (Arias, 1995), de forma que constituye el fundamento más trascendente de la teoría de muestreo.

Este método visto desde el punto de vista conceptual resulta el más sencillo. El mismo consiste en extraer todos los individuos al azar ubicados dentro del marco muestral o listado de valores otorgándoles a cada unidad igual probabilidad de inclusión, sería válido aclarar que la población debe ser relativamente pequeña o de estructura muy simple para garantizar que el diseño muestral se lleve a cabo de forma eficaz.

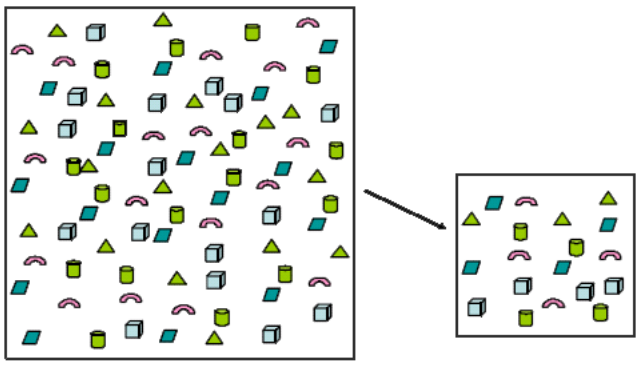


Figura 1.1: Representación gráfica del diseño muestral irrestricto aleatorio

ESTIMACIÓN DE UNA MEDIA Y UN TOTAL POBLACIONALES

Al usar muestreo irrestricto aleatorio para estimar la media poblacional, se obtiene el siguiente estimador:

$$\mu = \tilde{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Varianza Estimada de $\tilde{y}$:

$$\widehat{V}(\tilde{y}) = \frac{s^2}{n}(\frac{N-n}{N})$$

donde s^2 es la varianza muestral y se expresa por:

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \tilde{y})^2}{n-1}$$

Límite para el error de estimación:

$$2\sqrt{\widehat{V}(\tilde{y})} = 2\sqrt{(\frac{s^2}{n})(\frac{N-n}{N})}$$

Cotas para el error de estimación:

$$\tilde{y} \pm 2\sqrt{\widehat{V}(\tilde{y})}$$

El modelo aleatorio irrestrito permite además la inferencia de un estimador del total:

$$\widehat{\tau} = N\tilde{y} = N\frac{\sum_{i=1}^{n} y_i}{n}$$

Varianza Estimada de $\widehat{\tau}$:

$$\widehat{V}(\widehat{\tau}) = \widehat{V}(N\tilde{y}) = N^2(\frac{s^2}{n})(\frac{N-n}{N})$$

donde:

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \tilde{y})^2}{n-1}$$

Límite para el error de estimación:

$$2\sqrt{\widehat{V}(N\tilde{y})} = 2\sqrt{N^2(\frac{s^2}{n})(\frac{N-n}{N})}$$

Cotas para el error de estimación:

$$\widehat{\tau} \pm 2\sqrt{\widehat{V}(N\tilde{y})}$$

SELECCIÓN DEL TAMAÑO DE MUESTRA PARA LA ESTIMACIÓN DE LAS MEDIAS Y TOTALES POBLACIONALES

El tamaño de muestra requerido para estimar μ, con un límite para el error de estimación B:

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2}$$

donde:

$$D = \frac{B^2}{4}$$

El tamaño de muestra requerido para estimar τ, con un límite para el error de estimación B:

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2}$$

donde:

$$D = \frac{B^2}{4}N^2$$

ESTIMACIÓN DE UNA PROPORCIÓN POBLACIONAL

Estimación de la proporción poblacional para una muestra aleatoria simple:

$$\widehat{p} = \tilde{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Varianza Estimada de $\widehat{p}$:

$$\widehat{V}(\widehat{p}) = \frac{\widehat{p}\widehat{q}}{n-1}(\frac{N-n}{N})$$

donde:

$$\widehat{p} = 1 - \widehat{q}$$

Límite para el error de estimación:

$$2\sqrt{\widehat{V}(\widehat{p})} = 2\sqrt{\frac{\widehat{p}\widehat{q}}{n-1}(\frac{N-n}{N})}$$

Cotas para el error de estimación:

$$\widehat{p} \pm 2\sqrt{\widehat{V}(\widehat{p})}$$

SELECCIÓN DEL TAMAÑO DE MUESTRA PARA LA ESTIMACIÓN DE LA PROPORCIÓN POBLACIONAL

El tamaño de muestra requerido para estimar p, con un límite para el error de estimación B:

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2}$$

donde:

$$q = 1 - p \;\; y \;\; D = \frac{B^2}{4}$$

1.3.2. Muestreo Sistemático

El muestreo sistemático divide la población en n estratos seleccionándose un elemento; en el primero, el elemento se extrae bajo el procedimiento del método simple., pero, en los otros, queda determinado por la primera selección. El método de selección del muestreo sistemático se caracteriza por su sencillez y por el ahorro de tiempo que produce su utilización, lo simple de el trae como consecuencia la minimización de los errores de selección (Vinelo, 1990).

Este método de muestreo se identifica por particionar la población en subgrupos mediante del cálculo de una constante o coeficiente de elevación como resultado al dividir el tamaño del universo y el tamaño de la muestra. Luego se determina la primera extracción, pero, para esto, se debe elegir un número al azar o valor de arranque aleatorio comprendido entre 1 y coeficiente de elevación de ahí en adelante se tomará una unidad a intervalos regulares.

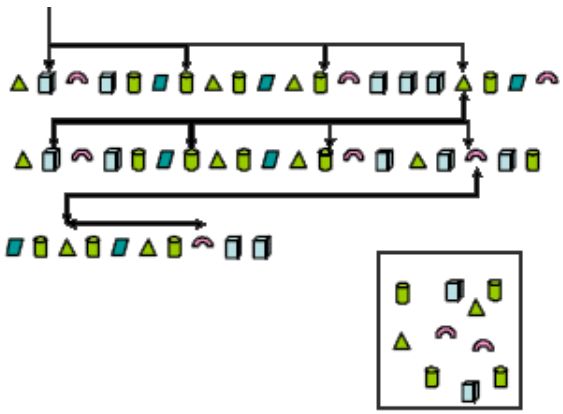


Figura 1.2: Representación gráfica del diseño muestral sistemático

ESTIMACIÓN DE UNA MEDIA Y UN TOTAL POBLACIONALES

Estimador de la media poblacional μ:

$$\widehat{\mu} = \tilde{y}_{sy} = \frac{\sum_{i=1}^{n} y_i}{n}$$

donde el subíndice sy significa que se utilizó el muestreo sistemático.

Varianza estimada de $\tilde{y}_{sy}$:

$$\widehat{V}(\tilde{y}_{sy}) = \frac{s^2}{n}(\frac{N-n}{N})$$

Límite para el error de estimación:

$$2\sqrt{\widehat{V}(\tilde{y}_{sy})} = 2\sqrt{(\frac{s^2}{n})(\frac{N-n}{N})}$$

Cotas para el error de estimación:

$$\tilde{y}_{sy} \pm 2\sqrt{\widehat{V}(\tilde{y}_{sy})}$$

Cuando el tamaño de la población es conocido se puede estimar $\widehat{\tau}$ mediante la siguiente expresión:

$$\widehat{\tau} = N\tilde{y}_{sy} = N\frac{\sum_{i=1}^{n} y_i}{n}$$

Varianza estimada de $\widehat{\tau}$:

$$\widehat{V}(N\tilde{y}_{sy}) = N^2\widehat{V}(\tilde{y}_{sy}) = N^2(\frac{s^2}{n})(\frac{N-n}{N})$$

Límite para el error de estimación:

$$2\sqrt{\widehat{V}(N\tilde{y}_{sy})} = 2\sqrt{N^2(\frac{s^2}{n})(\frac{N-n}{N})}$$

Cotas para el error de estimación:

$$\widehat{\tau} \pm 2\sqrt{\widehat{V}(N\tilde{y}_{sy})}$$

SELECCIÓN DEL TAMAÑO DE MUESTRA PARA ESTIMACIÓN DE MEDIAS Y PROPORCIONES POBLACIONALES

Tamaño de muestra requerido para estimar μ con un límite para el error de estimación:

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2}$$

donde:

$$D = \frac{B^2}{4}$$

Tamaño de muestra requerido para estimar p con un límite para el error de estimación:

$$n = \frac{Npq}{(N-1)D + pq}$$

donde:

$$q = 1 - p \;\; y \;\; D = \frac{B^2}{4}$$

1.3.3. Muestreo Aleatorio Estratificado

El muestreo aleatorio estratificado, consiste en subdividir la población en subpoblaciones de tal forma que la unión de ellas será la población y la intersección de cualquiera de dos dará como resultado el conjunto vacío, es decir, no tendrán elementos comunes. A las subpoblaciones se les llamará estratos se tratará de conformar estos de modo que los elementos dentro de ellos sean heterogéneos (Vinelo, 1990).

El diseño muestral aleatorio estratificado requiere de la estratificación de la población en subgrupos heterogéneos respecto a la característica en estudio, dichos subgrupos se les identificará como estratos. Su implementación parte de la distribución del tamaño muestral bajo el procedimiento irrestricto aleatorio a través de los diferentes estratos que conforman la población, este proceso se apoya en la utilización de *métodos de afijación del tamaño muestral* (Neyman, Proporcional, Óptimo).

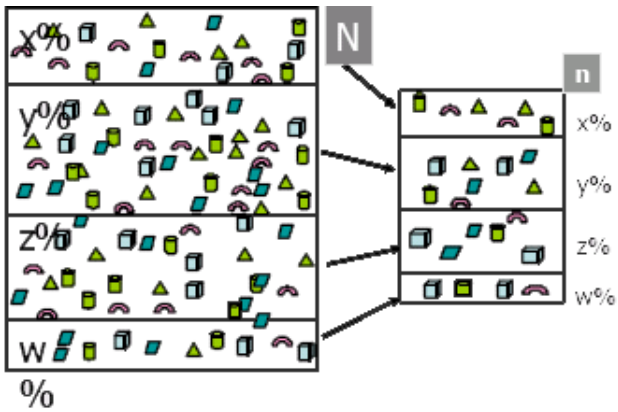


Figura 1.3: Representación gráfica del diseño muestral aleatorio estratificado

ESTIMACIÓN DE UNA MEDIA Y UN TOTAL POBLACIONALES

Estimador de la media poblacional μ:

$$\tilde{y}_{st} = \frac{1}{N} + [N_1\tilde{y}_1 + N_2\tilde{y}_2 + ... + N_L\tilde{y}_L] = \frac{1}{N}\sum_{i=1}^{L} N_i\tilde{y}_i$$

Varianza estimada de $\tilde{y}_{st}$:

$$\begin{aligned}\widehat{V}(\tilde{y}_{st}) &= \frac{1}{N^2}[N_1^2\widehat{V}(\tilde{y}_1) + N_2^2\widehat{V}(\tilde{y}_2) + ... + N_L^2\widehat{V}(\tilde{y}_L)] \\ &= \frac{1}{N^2}\sum_{i=1}^{L} N_i^2(\frac{N_i - n_i}{N_i})(\frac{s_i^2}{n_i})\end{aligned}$$

Límite para el error de estimación:

$$2\sqrt{\widehat{V}(\tilde{y}_{st})} = 2\sqrt{\frac{1}{N^2}\sum_{i=1}^{L} N_i^2(\frac{N_i - n_i}{N_i})(\frac{s_i^2}{n_i})}$$

donde:

$$s_i^2 = \frac{\sum_{j=1}^{n_i}(y_{ij} - \tilde{y}_i)^2}{n_i - 1}$$

Cotas para el error de estimación:

$$\tilde{y}_{st} \pm 2\sqrt{\widehat{V}(N\tilde{y}_{st})}$$

La estimación correspondiente a un total poblacional τ viene dado por la siguiente expresión:

$$N\tilde{y}_{st} = N_1\tilde{y}_1 + N_2\tilde{y}_2 + ... + N_L\tilde{y}_L = \sum_{i=1}^{L} N_i\tilde{y}_i$$

Varianza estimada de $N\tilde{y}_{st}$:

$$\widehat{V}(N\tilde{y}_{st}) = N^2\widehat{V}(\tilde{y}_{st}) = \sum_{i=1}^{L} N^2(\frac{N_i - n_i}{N_i})(\frac{{s_i}^2}{n_i})$$

Límite para el error de estimación:

$$2\sqrt{\widehat{V}(N\tilde{y}_{st})} = 2\sqrt{\sum_{i=1}^{L} N^2(\frac{N_i - n_i}{N_i})(\frac{{s_i}^2}{n_i})}$$

Cotas para el error de estimación:

$$\widehat{\tau} \pm 2\sqrt{\widehat{V}(N\tilde{y}_{st})}$$

SELECCIÓN DEL TAMAÑO DE MUESTRA PARA LA ESTIMACIÓN DE MEDIAS Y TOTALES POBLACIONALES

La cantidad de información presente en una muestra está determinada por el número de unidades que la forman, esto se debe a que la $\widehat{V}(\tilde{y}_{st})$ decrece con el aumento de n. Una vez determinado el tamaño muestral en el diseño Aleatorio Estratificado surge la interrogante de cómo dividir n entre los tamaños tamaños de muestra de los estratos $n_1, n_2, ..., n_L$. Con la finalidad de resolver este dilema se pretende la utilización de métodos de asignación de la muestra que presente una cantidad de información especificada a un costo mínimo.

ASIGNACIÓN ÓPTIMA DEL TAMAÑO DE MUESTRA

Asignación aproximada que minimiza el costo para el valor fijo de $\widehat{V}(\tilde{y}_{st})$ o que minimiza $\widehat{V}(\tilde{y}_{st})$ para un costo fijo:

$$\begin{aligned} n_i &= n\left(\frac{N_i\sigma_i/\sqrt{c}_i}{N_1\sigma_1/\sqrt{c}_1 + N_2\sigma_2/\sqrt{c}_2 + ... + N_L/\sqrt{c}_L}\right) \\ &= n\left(\frac{N_i\sigma_i/\sqrt{c}_i}{\sum_{k=1}^{L} N_k\sigma_k/\sqrt{c}_k}\right) \end{aligned}$$

donde N_i denota el tamaño de i-ésimo estrato, σ_i la varianza poblacional para el i-ésimo estrato y c_i denota el costo asociado para obtener una observación individual del i-ésimo estrato. luego la determinación del tamaño muestral vendrá dado por la siguiente expresión:

$$n = \frac{(\sum_{k=1}^{L} N_k\sigma_k/\sqrt{c}_k)(\sum_{i=1}^{L} N_i\sigma_i\sqrt{c}_i)}{N^2D + \sum_{i=1}^{L} N_i\sigma_i^2}$$

para asignación óptima con la varianza de $\tilde{y}_{st}$ fija en D.

ASIGNACIÓN DE NEYMAN DEL TAMAÑO DE MUESTRA

En ciertos problemas que involucran el diseño muestreo estratificado, los costes relacionados para la obtención de información pueden resultar los mismos para todos los estratos que conforman la población o en cambio pueden ser considerados como desconocidos. Sería razonable entonces, suponer que los costos son iguales, es decir, $c_1 = c_2 = ... = c_L$, luego la asignación de la muestra está dado por la siguiente expresión:

$$n_i = n\left(\frac{N_i\sigma_i}{\sum_{i=1}^{L} N_i\sigma_i}\right)$$

entonces para el tamaño de muestra total n la expresión toma la forma:

$$n = \left(\frac{\sum_{i=1}^{L} N_i\sigma_i^2}{N^2D + \sum_{i=1}^{L} N_i\sigma_i^2}\right)$$

ASIGNACIÓN PROPORCIONAL DEL TAMAÑO DE MUESTRA

Además de encontrar valores de costos iguales, en algunas ocaciones resultan las mismas varianzas, $\sigma_1^2, \sigma_2^2, ...\sigma_L^2$. Para este caso la asignación de la muestra se realiza de la siguiente forma:

$$n_i = n(\frac{N_i}{\sum_{i=1}^{L} N_i})$$

luego, en asignación proporcional la determinación del valor de n pasa a ser:

$$n = (\frac{\sum_{i=1}^{L} N_i \sigma_i^2}{ND + \frac{1}{N}\sum_{i=1}^{L} N_i \sigma_i^2})$$

ESTIMACIÓN DE UNA PROPORCIÓN POBLACIONAL

Estimador de la proporción poblacional p:

$$\widehat{p}_{st} = \frac{1}{N}(N_1\widehat{p}_1 + N_2\widehat{p}_2 + ... + N_L\widehat{p}_L) = \frac{1}{N}\sum_{i=1}^{L} N_i\widehat{p}_i$$

Varianza estimada de $\widehat{p}_{st}$:

$$\begin{aligned} \widehat{V}(\widehat{p}_{st}) &= \frac{1}{N^2}[N_1^2\widehat{V}(\widehat{p}_1) + N_2^2\widehat{V}(\widehat{p}_2) + ... + N_L^2\widehat{V}(\widehat{p}_L)] \\ &= \frac{1}{N^2}\sum_{i=1}^{L} N_i^2(\frac{N_i - n_i}{N_i})(\frac{\widehat{p}_i\widehat{q}_i}{n_i - 1}) \end{aligned}$$

Límite para el error de estimación:

$$2\sqrt{\widehat{V}(\widehat{p}_{st})} = 2\sqrt{\frac{1}{N^2}\sum_{i=1}^{L} N_i^2(\frac{N_i - n_i}{N_i})(\frac{\widehat{p}_i\widehat{q}_i}{n_i - 1})}$$

Cotas para el error de estación:

$$\widehat{p}_{st} \pm 2\sqrt{\widehat{V}(\widehat{p}_{st})}$$

SELECCIÓN DEL TAMAÑO DE MUESTRA Y ASIGNACIÓN DE LA MUESTRA PARA LA ESTIMACIÓN DE PROPORCIONES

Para estimar una proporción poblacional, se debe partir del establecimiento del error máximo permisible, esto posibilitará la elección de un tamaño muestral que proporcione el nivel de información deseado por el experimentador.

ASIGNACIÓN ÓPTIMA DEL TAMAÑO DE MUESTRA

Asignación aproximada que minimiza el costo para el valor fijo de $V(\widehat{p}_{st})$ para un costo fijo:

$$\begin{aligned} n_i &= n\frac{N_i\sqrt{p_iq_i/c_i}}{N_1\sqrt{p_1q_1} + N_2\sqrt{p_2q_2} + ... + N_L\sqrt{p_Lq_L}} \\ &= n\frac{N_i\sqrt{p_iq_i/c_i}}{\sum_{k=1}^{L}\sqrt{p_kq_k/c_k}} \end{aligned}$$

donde N_i denota el tamaño del i-ésimo estrato, p_i denota la proporción poblacional para el i-ésimo estrato y c_i denota el costo por obtener una observación individual del i-ésimo estrato.

Tamaño de muestra aproximado que se requiere para estimar p, con un límite B para el error de estimación:

$$n = \frac{\sum_{i=1}^{L} N_i^2 p_i q_i / w_i}{N^2 D + \sum_{i=1}^{L} N_i p_i q_i}$$

donde w_i es la fracción de observaciones asignada al estrato i,p_i es la proporción poblacional del estrato i, y donde:

$$D = \frac{B^2}{4}$$

ASIGNACIÓN DE NEYMAN DEL TAMAÑO DE MUESTRA

Si el costo de muestreo no varía de un estrato a otro, entonces los factores de costo son ignorados en la expresión de asignación del tamaño muestral.

$$n_i = n(\frac{N_i \sqrt{\widehat{p_i}\widehat{q_i}}}{\sum_{i=1}^{L} N \sqrt{\widehat{p_i}\widehat{q_i}}})$$

luego la expresión del tamaño muestral nos queda :

$$n = \frac{(\sum_{i=1}^{L} N \sqrt{\widehat{p_i}\widehat{q_i}})^2}{N^2 D + \sum_{i=1}^{L} N \widehat{p_i}\widehat{q_i}}$$

ASIGNACIÓN PROPORCIONAL DEL TAMAÑO DE MUESTRA

Al efectuar la asignación del tamaño muestral puede darse el caso de que las varianzas, así como los costos sean iguales en todos los estratos, realizando una analogía respecto al valor de la proporción p_i correspondiente a cada estrato se deberá emplear la siguiente expresión:

$$n_i = n(\frac{N}{N})$$

donde el tamaño de muestra n se puede determinar usando la expresión:

$$n = \frac{\sum_{i=1}^{L} N_i p_i q_i}{N D + \sum_{i=1}^{L} N_i p_i q_i}$$

1.3.4. Muestreo Por Conglomerado Monoetápico

En este modelo basado en particiones, las clases reciben la denominación general de conglomerados, si bien cuando la definición se establece en términos espaciales (usualmente geográficos: municipios, distritos, barrios,...), suele hablarse de muestreo de áreas.

En primer lugar, se establece una partición N que genera n clases. En un segundo momento, se establece la selección de las clases, no de las unidades elementales que la componen, la muestra se construye pues con la unión de las clases seleccionadas (Arrondo, 1996).

Para (Mendenhall, 1987), el muestreo por conglomerados es un diseño efectivo para obtener una cantidad especificada de información al costo mínimo bajo las siguientes condiciones:

- No se encuentra disponible o es muy costoso obtener un buen marco que liste los elementos de la población, mientras que se puede lograr fácilmente una marco que los liste los conglomerados.
- El costo por obtener observaciones se incrementa con la distancia que separa los elementos.

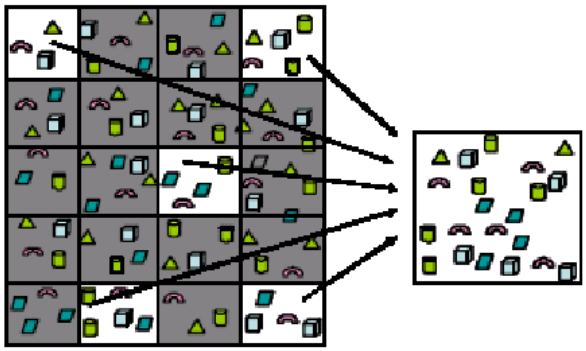


Figura 1.4: Representación gráfica del diseño muestral por conglomerado monoetápico

Para lograr una mejor compresión de esta técnica de muestreo se establece la siguiente notación:

$N =$ número de conglomerados en la población

$n =$ número de conglomerados seleccionados en una muestra

$m_i =$ número de elementos en el conglomerado $i, i = 1, ...N$

$\tilde{m} = \frac{1}{n}\sum_{i=1}^{N} m_i =$ tamaño promedio del conglomerado en la muestra

$M = \sum_{i=1}^{N} m_i =$ número de elementos en la población

$\tilde{M} = \sum_{i=1}^{N} \frac{M}{N}$ = tamaño promedio del conglomerado en la población

y_i = total de todas las observaciones en el i-ésimo conglomerado

ESTIMACIÓN DE UNA MEDIA Y UN TOTAL POBLACIONALES

Estimador de la media poblacional μ:

$$\tilde{y} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i}$$

Varianza estimada de $\tilde{y}$:

$$\widehat{V}(\tilde{y}) = (\frac{N-n}{Nn\tilde{M}^2})\frac{\sum_{i=1}^{n}(y_i - \tilde{y}m_i)^2}{n-1}$$

Límite para el error de estimación:

$$2\sqrt{\widehat{V}(\tilde{y})} = 2\sqrt{(\frac{N-n}{Nn\tilde{M}^2})\frac{\sum_{i=1}^{n}(y_i - \tilde{y}m_i)^2}{n-1}}$$

Cotas para el error de estimación:

$$\tilde{y} \pm 2\sqrt{\widehat{V}(\tilde{y})}$$

Para estimar el total poblacional τ se tendrá en cuenta de forma análoga al muestreo irrestricto aleatorio $M\tilde{\mu}$ lo cual implica que $M\tilde{y}$ proporcione un estimador de μ.

Estimador del total poblacional τ:

$$M\tilde{y} = M\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i}$$

Varianza estimada de $M\tilde{y}$:

$$\widehat{V}(M\tilde{y}) = M^2\widehat{V}(\tilde{y}) = N^2(\frac{N-n}{Nn})\frac{\sum_{i=1}^{n}(y_i - \tilde{y}m_i)^2}{n-1}$$

Límite para el error de estimador:

$$2\sqrt{\widehat{V}(M\tilde{y})} = 2\sqrt{N^2(\frac{N-n}{Nn})\frac{\sum_{i=1}^{n}(y_i - \tilde{y}m_i)^2}{n-1}}$$

Cotas para el error de estimación:

$$\tau \pm 2\sqrt{\widehat{V}(M\tilde{y})}$$

ESTIMACIÓN DE UNA PROPORCIÓN POBLACIONAL

Para estimar la proporción poblacional p se considera como su mejor estimador la proporción muestral $\widehat{p}$. Sea a_i el número total de elementos en el conglomerado i que poseen la característica

de interés, entonces la proporción de elementos en la muestra de n conglomerados es dada por:

Estimador de la proporción poblacional p:

$$\widehat{p} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i}$$

Varianza estimada de $\widehat{p}$:

$$\widehat{V}(\widehat{p}) = (\frac{N-n}{Nn\tilde{M}^2})\frac{\sum_{i=1}^{n}(a_i - \widehat{p}m_i)^2}{n-1}$$

Límite para el error de estimación:

$$2\sqrt{\widehat{V}(\widehat{p})} = 2\sqrt{(\frac{N-n}{Nn\tilde{M}^2})\frac{\sum_{i=1}^{n}(a_i - \widehat{p}m_i)^2}{n-1}}$$

Cotas para el error de estimación:

$$\widehat{p} \pm 2\sqrt{\widehat{V}(\widehat{p})}$$

SELECCIÓN DEL TAMAÑO DE MUESTRA PARA LA ESTIMACIÓN DE MEDIAS Y TOTALES POBLACIONALES

Tamaño de muestra aproximado para estimar μ con un límite B para el error de estimación:

$$n = \frac{N\sigma_c^2}{ND + \sigma_c^2}$$

donde σ_c^2 es estimado por s_c^2:

$$s_c^2 = \frac{\sum_{i=1}^{n}(y_i - \tilde{y}m_i)^2}{n-1}$$

y

$$D = \frac{B^2\tilde{M}^2}{4}$$

Tamaño de muestra requerido para estimar τ con un límite B para el error de estimación:

$$n = \frac{N\sigma_t^2}{ND + \sigma_t^2}$$

donde σ_t^2 se estima mediante s_t^2

$$s_t^2 = \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_t)^2}{n-1}$$

y

$$D = \frac{B^2}{4N^2}$$

1.3.5. Muestreo Piloto

El estudio piloto o pre-test es un ***"diseño muestral iterativo"*** concebido para la determinación del número de elementos que conforman el tamaño muestral. La razón principal que origina la puesta en práctica de este procedimiento es el desconocimiento del valor de la varianza o proporción (puede conocerse de estudios anteriores) de la población en estudio para la posterior estimación del tamaño de muestra correspondiente. A continuación se describe el procedimiento a seguir:

1. Se extrae una muestra pequeña de tamaño M la cual se requiere inicialmente para un cálculo preliminar de la varianza, por ello se plantea la toma de una muestra pequeña.

2. Se calcula la varianza a partir de la muestra tomada de tamaño M y se determina el tamaño muestral n a través de la fórmula:

$$n = \frac{S^2{}_M t^2{}_{\frac{\alpha}{2}}; M-1}{d}$$

3. Verificar que la nueva muestra cumple con los siguientes requerimientos

   a) Coeficiente de confianza 1-α

   b) Ancho del intervalo 2d

4. Si el intervalo resulta más ancho que dos veces el error d se volverá a calcular n pero considerando como muestreo piloto toda la muestra extraída y con esto se logra aproximarse secuencialmente hasta lograr el intervalo deseado.

De igual forma esta problemática aparece cuando se estudia el comportamiento de cierta característica de los elementos en la población y se desea estimar su proporción respecto a algún criterio preestablecido. Esto se soluciona efectuando aproximaciones sucesivas de manera que se logre el cumplimiento de determinados intervalos. Los pasos a seguir son los siguientes:

1. Se toma una pequeña muestra que en este caso no puede ser menor que 30 unidades de acuerdo a la tabla de Cochran ver **Anexo 1**. Se calcula el valor de la proporción $\widehat{p}$ en esta muestra y se sustituye en la fórmula:

$$n = (\frac{Z_{\frac{\alpha}{2}}}{d})^2[\widehat{p}_M(1-\widehat{p}_M)]$$

   donde se observa la sustitución de p por p_M.

2. Se toma una muestra del tamaño calculado y se obtiene un nuevo valor de proporción $\widehat{p}$. Si se cumplen las siguientes dos restricciones:

   - $$n = (\frac{Z_{\frac{\alpha}{2}}}{d})^2[\widehat{p}(1-\widehat{p})]$$
   - n cumple las recomendaciones de la tabla de Cochran para el valor de proporción $\widehat{p}$ calculado.

Se puede asegurar que el valor de muestra calculado será el necesario. Si se incumple la primera restricción se repite el procedimiento partiendo del nuevo valor de n teniéndose en cuenta como $\widehat{p}_M$ el valor $\widehat{p}$ obtenido con la segunda muestra. Este procedimiento se detendrá al cumplir ambas restricciones. Si en caso contrario se cumple la restricción primera y no siendo así la segunda se seleccionará el tamaño muestral recomendado por la tabla de Cochran.

1.4. Descripción del objeto de estudio

1.4.1. Flujo actual de los procesos y análisis crítico de la ejecución de estos

En la actualidad, el proceso concerniente al desarrollo de una investigación mediante el diseño de encuestas por muestreo, se desempeña de forma manual con el apoyo de tablas de números aleatorios, dígase, la confección de reportes relacionados con la determinación del tamaño de muestra y los elementos seleccionados para realizar la inferencia de los valores de la población. En el análisis posterior de los datos en muchas ocasiones se introduce la utilización de softwares tales como el SPSS o el Excel que ofrecen opciones gráficas entre las cuales se encuentran las tablas de frecuencia, histogramas etc.

El investigador y/o grupo investigativo se encargan de establecer los diferentes niveles organizativos y de dirección, además, la posterior capacitación de todo el personal involucrado en el proceso investigativo, tales como entrevistadores, enumeradores, el personal destinado al procesamiento de los datos obtenidos, ello permitirá desempeñar a cabalidad el trabajo de campo al cual indudablemente se le concierne gran importancia.

A pesar de una adecuada labor por parte de todos los participantes en la puesta en práctica del modelo de encuesta, en la actualidad, muchas veces lograr la viabilidad y confiabilidad de la muestra, así como, en los estadísticos obtenidos a partir de la misma, resulta en efecto tedioso y complicado sin la presencia de un software. En ocasiones se hace necesario implementar un estudio preliminar o prueba piloto para la eliminación de posibles errores en la elaboración de la encuesta, de no contarse con un software capaz de guiar al investigador durante su desarrollo (iterativo) se incurriría en una considerable pérdida de tiempo y recursos. Resultados como conocer las unidades seleccionadas en la muestra, así como las unidades asignadas por conglomerados, estratos, implican una compleja confección manualmente, ello naturalmente, se tornaría fácil mediante el empleo de un sistema informático.

1.5. Descripción de los sistemas existentes

1.5.1. Algunos software estadísticos a nivel internacional

Existen a nivel internacional variados software de propósito general entorno a la estadística que abordan el problema del muestreo, entre ellos se encuentran:

- **MINITAB**: Ofrece comandos relacionados con la generación de datos de distribuciones usuales mediante una muestra aleatoria de cualquier columna para la hoja de datos o de una distribución de probabilidad dada. Otras de las opciones de que dispone este software está relacionado con la determinación de la función empírica asociada a una muestra, así como, realizar la aproximación a la distribución en el muestreo (*Método de Montecarlo*).

  MINITAB contiene un lenguaje de programación sencillo pero potente, que permite elaborar una gran variedad de programas hechos a la medida del usuario. Estos programas se llaman macros. Las instrucciones de las macros pueden contener los típicos controladores de flujo que se usan en los lenguajes de programación.

  Un elemento que delimita el uso del presente software en relación al muestreo viene dado en que, el mismo, está desprovisto de los diseños muestrales necesarios para el estudio los estadísticos poblacionales.

- **EXCEL**: Software incluido en el paquete de programas Microsoft Office, dispone de opciones relacionadas con el desempeño del cálculo de la teoría del muestreo referentes a la estimación de medias, varianzas, totales, intervalos de confianza etc., pero el mismo carece de un módulo específico para el cálculo de tamaños de muestra.

  Para la llevar a cabo la determinación del número de elementos de la muestra se apoya en la programación de las fórmulas según el diseño muestral en cuestión mediante el uso de la hoja de cálculo, ello sugiere un conocimiento profundo de la temática por parte del experimentador.

- **STATGRAPHICS**: Este software propone un módulo referente a la determinación de tamaños de muestras relacionado al uso de parámetros binomiales, normal sigma, proporción binomial, y la distribución de Poisson, además dispone de tres módulos relacionados con la teoría del muestreo tales como:

  - Análisis de una muestra
  - Comparación de dos muestras
  - Comparación de múltiples muestras

  De manera general el enfoque propuesto en esta aplicación no contempla ninguno de los diseños del muestreo indispensables para una adecuada concepción del proceso de encuesta por muestreo.

- **SPSS v15**: Dispone con un módulo Muestras Complejas módulo add-on de SPSS para Windows que proporciona herramientas especializadas de planificación y los estadísticos necesarias para trabajar con datos muestrales. Permite realizar inferencias estadísticas de una población al incorporar el diseño de la muestra al análisis. Muestras Complejas trabaja con diseños muestrales, tales como diseños estratificados, de conglomerados o muestreos polietápicos, necesita técnicas especiales para considerar en el diseño de muestreo los errores estándares asociados.

  Es válido señalar que este software no contempla el diseño muestral piloto como herramienta auxiliar para corroborar posibles errores en la elaboración del diseño de encuesta y calibrar el tamaño de muestra mínimo requerido para estimar de forma preliminar los estadísticos de la población.

1.5.2. Algunos software estadísticos existentes en Cuba

Después de realizar una investigación detallada sobre la existencia de algún software destinado para el cálculo del tamaño de muestras enmarcado en nuestro país, se obtuvo como resultado la existencia del software MUESTRAUD (Muestreo Auditoría).

El desarrollo de este programa va dedicado a la actividad auditora en la Empresa Termoeléctrica Cienfuegos ofrece la posibilidad de seleccionar los elementos muestrales, según las técnicas de muestreo aplicadas para la verificación y el control de los artículos en inventario, a saber el Muestreo Aleatorio Simple (MAS), el Muestreo Sistemático (MS) y Muestreo Aleatorio Estratificado (MAE).

Las principales deficiencias que presenta este software se centran en la carencia del diseño muestral por Conglomerado en una Etapa como una técnica de importancia relevante en el desarrollo de cualquier investigación de índole muestral, a ello se suma la no disponibilidad de comandos específicos para la puesta en practica del diseño muestral piloto.

1.5.3. Presentación de la propuesta de solución

La propuesta de solución lleva por nombre **SICTM S**istema **I**nformático para el **C**álculo de la **T**eoría del **M**uestreo, tomando como punto de partida los diferentes diseños muestrales o técnicas de muestreo previstos en su concepción.

SICTM contribuye a la reducción de los costos partiendo del concepto de minimizar el número de unidades a seleccionar en relación al universo de análisis, además se obtiene una mayor información por unidad de costo, resulta menor el tiempo que se necesita para la recogida de la información, su procesamiento, puesto que el número de elementos necesarios en la enumeración es menor que si fuese sobre toda la población, agiliza la gestión del trabajo relacionado con las tablas de números aleatorios y por ende los reportes asociados al tamaño muestral y las unidades que lo conforman.

El presente producto de software a diferencia de los descritos con anterioridad permite mejorar notablemente el proceso investigativo, facilitándole al usuario la determinación de tamaños de muestras de forma sencilla y rápida sin disminuir la confiabilidad en los resultados, basándose para esto, en la utilización de modelos matemáticos.

El sistema posibilita el almacenamiento e impresión de la información contenida tanto en la determinación del número de elementos de la muestra, como la estimación de los parámetros poblacionales a partir de su inferencia estadística.

1.6. Tendencias, metodologías y/o tecnologías actuales

Durante el proceso de desarrollo de un software e independientemente de su procedencia científica se deben tener en cuenta determinados aspectos encaminados a garantizar eficiencia y calidad. Por esta razón se hace necesario prescindir de las diferentes metodologías y en correspondencia a nuestra situación problemática hacer uso consciente de alguna de estas.
Otro elemento a tener en cuenta es la selección de las tecnologías de impacto que se ajusten adecuadamente a las necesidades del producto informático, para esto se debe realizar con anterioridad un estudio detallado de las mismas, analizando las diferentes ventajas y desventajas en su puesta en práctica con la finalidad de emplear las más convenientes.

1.6.1. Fundamentación de la metodología utilizada

Lenguaje de Modelamiento Unificado (UML)

Lenguaje Unificado de Modelado (UML, por sus siglas en inglés, Unified Modeling Language) es el lenguaje de modelado de sistemas de software más conocido y utilizado en la actualidad; aún cuando todavía no es un estándar oficial, está respaldado por el OMG (Object Management Group). UML no es un lenguaje de programación sino un lenguaje de propósito general para el modelado orientado a objetos y también puede considerarse como un lenguaje de modelación visual que permite una abstracción del sistema y sus componentes (?). Es un lenguaje gráfico para visualizar, especificar, construir y documentar un sistema de software. Ofrece un estándar para describir un ”plano”del sistema (modelo), incluyendo aspectos conceptuales tales como procesos de negocios y funciones del sistema, y aspectos concretos como expresiones de lenguajes de programación, esquemas de bases de datos y componentes de software reutilizables.

Es importante resaltar que UML es un ”lenguaje”para especificar y no para describir métodos o procesos. Se utiliza para definir un sistema de software, para detallar los artefactos en el

sistema y para documentar y construir. En otras palabras, es el lenguaje en el que está descrito el modelo. Se puede aplicar en una gran variedad de formas para dar soporte a una metodología de desarrollo de software (tal como el Proceso Unificado Racional), pero no especifica en sí mismo qué metodología o proceso usar.

Análisis y Diseño Orientado a Objetos de Sistemas Informáticos (ADOOSI-UML) versión 5.0
La metodología ADOOSI fue creada en el año 2000 en el Centro de Estudios de Ingeniería y Sistemas adscrito al ISPJAE, en la Habana, por las doctoras Sofía Álvarez y Anaisa Hernández como respuesta a la necesidad de poseer un conjunto de métodos e instrumentos que se unieran de forma tal que permitieran el desarrollo lógico de proyectos que utilicen el enfoque orientado a objetos (OO) siguiendo cada una de las etapas que constituyen su ciclo de vida. En esta metodología, a diferencia de la forma de trabajo tradicional, el análisis, el diseño y el desarrollo se realizan de forma concurrente por lo que las etapas no coinciden con las establecidas en otras metodologías orientadas a objetos existentes.
En su versión actual se aprovechan las facilidades que brindan los medios ambientes visuales, el auge de la técnica de prototipación que brindan los medios ambientales visuales, la extensión hacia OO del modelo relacional, el desarrollo de lenguajes de programación OO que trabajen con bases de datos relacionales y la aceptación como estándar de la notación UML (Unified Modeling Language).

1.6.2. Fundamentación del lenguaje y software utilizado

Borland Delphi 7
Delphi, desciende por decirlo de alguna manera, del mítico lenguaje desarrollado por Niklaus Wirth en 1971, para facilitar la enseñanza de la programación estructurada, y que le puso el nombre del tan celebre filósofo y matemático francés Pascal. En aquella época, las fases de edición de código, compilación y enlazado, iban por separado, hasta que Philippe Khan, el fundador de Borland, actualmente Inprise, creo el Turbo Pascal, el primer entorno de programación DOS, con el editor, compilador y linkador integrados, siguiéndole los famosos Turbo Basic y Turbo C/C++.
Características:

- Herramienta de desarrollo más rápida y productiva para la web y desarrollo de Bases de Datos. No hay que olvidar que pronto, gracias al comercio electrónico, la gestión irá ligada a

la web.

- Velocidad de ejecución, compilación y enlace cercanas al C++, y por lo tanto mucho mejores que otros lenguajes existentes.
- Programación Orientada a Objetos verdadera, permite encapsulamiento,herencia y polimorfismo.
- Componente integrados dentro del lenguaje, lo que reduce considerablemente la utilización de librerías y controles externos, por lo tanto menos problemas para nosotros y el usuario.
- Tratamiento de errores mediante excepciones, lo que impide el típico error de programa que nos echa fuera.
- Soporte avanzado de Bases de Datos mediante BDE (Borland Database Engine), ADO (ActiveX Database Objects), tecnología de Microsoft de acceso a Bases de Datos e Internet y finalmente InterBase Express, acceso nativo a InterBase, para desarrollo Cliente/Servidor off-line.
- Modelo de datos y relaciones de forma visual.
- Asistentes y componentes para Internet/Intranet.
- Componentes compatibles con Microsoft Office.
- Fácil integración de informes y gráficos de gestión.
- Facilidad en la distribución de aplicaciones con el InstallShield Express.

Borland Builder C++ versión 6

Borland C++Builder 6 Standard constituye la vía más rápida para desarrollar aplicaciones C++.
Características:

- C++Builder 6 ofrece el entorno de desarrollo visual que miles de desarrolladores C++ buscan a la hora de crear sus aplicaciones. Desarrolle rápidamente prototipos y aplicaciones completas, mediante una amplia paleta de componentes con más de 85 componentes reutilizables.

- Los conocimientos de programación quedan protegidos gracias al uso de código 100 % estándar ANSI/ISO, sin extensiones propietarias que limiten la elección de futuras herramientas de desarrollo. Si dispone de código C++ existente, puede importarlo directamente en C++Builder.

- Posibilita la gestión eficaz de proyectos mediante la herramienta Advanced Project Manager, para el control de los archivos fuentes utilizados en su proyecto, ello hace posible visualizar las aplicaciones y las distribuciones realizadas.

- Muchas de las herramientas que forman parte del entorno de desarrollo integrado han sido mejoradas para incrementar su productividad, incluyendo muchos Asistentes para la mayoría de tareas habituales.

- El compilador C++ incluido, Borland C++ Compiler 5.5, es un compilador y optimizador del código, de alto rendimiento y multihebrado, que actúa en segundo plano. Sin dejar de trabajar, sus aplicaciones se compilarán y ejecutarán más rápidamente.

1.7. Conclusiones Parciales

En este capítulo, después de desarrollar un estudio profundo de los aspectos teóricos en relación al tema, se demostró que la elaboración del sistema informático para la determinación del número de elementos de la muestra constituye una herramienta auxiliar en la confección del modelo de encuesta, debido a la gestión del tamaño muestral a partir de modelos matemáticos que garantizan la confiabilidad y representatividad de los datos que la contienen. Así mismo, permite acelerar el tiempo requerido para el procesamiento de la información recogida de la muestra y por consiguiente se minimizan los costos asociados al desarrollo de la investigación.

El análisis realizado sobre las diferentes metodologías y tecnologías existentes en la actualidad ha permitido considerar RUP como metodología a seguir para la documentación del sistema en su proceso de desarrollo. Además, como lenguaje de programación se ha seleccionado el C++ y como herramienta de desarrollo el Borland C++ Builder v6.

Capítulo 2

Descripción y construcción de la solución propuesta

2.1. Introducción

En este capítulo, con la finalidad de lograr una mejor comprensión de las diferentes actividades que se desempeñan durante el proceso de investigación científico se exponen los diferentes conceptos y entidades involucrados en la situación problémica, las reglas del negocio a considerar, así como, el diagrama de clases del modelo de objetos del dominio en el cual se a reflejado el modo de operación del sistema informático. Con el propósito de alcanzar la correcta modelación e implementación de la solución propuesta se ha empleado la metodología RUP, determinándose los requerimientos funcionales y no funcionales, de igual forma se definen los actores y casos de uso correspondientes al sistema, vistos a través del diagrama de casos de uso del sistema con la descripción de cada uno de los mismos. Se presentan los diagramas de clases del diseño y de implementación correspondientes al sistema. Se mencionan los aspectos relacionados con el diseño de la aplicación, además se explica el tratamiento de los errores y la concepción de la ayuda del sistema.

2.2. Descripción del modelo del dominio

2.2.1. Definición de entidades y conceptos principales

El modelado del dominio es utilizado para definir el modelado sistemático de funciones, los objetos, los datos y las relaciones de los sistemas de software. Esto, posibilitará un conocimiento de las características del sistema de software empleado, un vocabulario común que capture el modelo en que las personas interesadas comprendan el dominio.
El proceso de selección del tamaño muestral mediante la encuesta por muestreo comprende una gama de conceptos básicos de importancia relevante, entre los cuales se encuentran los siguientes:

La idea fundamental cuando se pretende llevar a cabo el estudio de una población parte del adecuado establecimiento de la misma tomando como referencia la(s) característica(s) de interés de la investigación. Por lo tanto, es necesario definir la ***población objetivo*** como un conjunto de medidas, o el recuento de todas las unidades que presentan una característica en común, se podría definir como un conjunto de mediciones finito o infinito, real o conceptual.
El objetivo de la teoría del muestreo es la estimación de los estadísticos de una población a partir de la información contenida en una muestra, luego, para dicho tamaño de muestra se establece el *marco de investigación* constituido por un listado actualizado de todos los elementos que conforman la población objeto de investigación, también puede ser un mapa o croquis con las unidades de selección plenamente identificados.
De manera general el *elemento o unidad de selección* puede ser una persona, familia, empresa, zona, animal u objeto, del elemento se estudian sus características. Estas se clasifican en:

- cualitativas a atributos expresadas por palabras, y se cuantifican mediante el conteo o recuento.
- cuantitativas o variables expresadas en forma numérica, que pueden ser medidas (variables continuas) o contadas (variables discretas).

Una investigación de índole muestral puede ser abordada por el experimentador a través de dos puntos de vista diferentes, es decir, como una primera alternativa los ***diseños de muestreo probabilísticos*** donde su característica fundamental es que todo elemento del universo tiene asociado una determinada probabilidad de integrar la muestra, dicho valor de probabilidad puede

ser expresado a través de modelos matemáticos.
Los procedimientos más usuales para la obtención de muestras aleatorias son los siguientes:

- Muestreo Irrestricto Aleatorio
- Muestreo Aleatorio Estratificado
- Muestreo Sistemático
- Muestreo Por Conglomerados Monoetápico
- Muestreo Por Conglomerados Bietápico

Una segunda opción la sería los ***diseños muestrales no probabilísticos*** donde el experimentador procede en cierta forma a ciegas, pues no tiene idea del error que puede estar introduciendo en sus apreciaciones. Las muestras no probabilísticas más utilizadas son las llamadas *accidentales*, por *cuotas o intencionales*.

- ***Accidentales:*** Tamaño muestral que se obtiene sin ningún plan preconcebido, resultando las unidades que la forman producto de circunstancias fortuitas.
- ***Por Cuotas:*** Consiste en predeterminar la cantidad de elementos de cada categoría que conformarán la muestra de forma arbitraria por lo que la rigurosidad de la muestra se reduce considerablemente.
- ***Intencional:*** Escoge sus unidades de forma completamente arbitraria, designando a cada unidad según características que para el investigador resulten de relevancia. Estas muestras son muy útiles y se emplean frecuentemente en los estudios de caso, por más que la posibilidad de generalizar conclusiones, a partir de ellas, sea en rigor nula.

Una vez establecido el tamaño de muestra que cumple los requerimientos propuestos y efectuada la recolección de los datos obtenidos al aplicar la encuesta el experimenador está en condiciones de inferir los ***parámetros poblacionales*** estos no son más que las medidas descriptivas numéricas aplicadas a las características de las unidades de la población en estudio, denominadas como valores estadísticos de la población. La *estimación por intervalos* es la estimación del parámetro mediante la especificación de un intervalo de valores, determinada por un limite inferior y otro superior (límites de confianza) dentro del cual estará comprendido el valor verdadero o parámetro

poblacional.

Se considera que un buen estimador debe ser:

- ***Insesgado:*** Es decir que no tenga sesgo, error o bias, cuando el error del estimador es igual al del parámetro. En caso contrario la estimación será sesgada.
- ***Consistente:*** Es aquel estimador que al aumentar el tamaño de la muestra, converge en probabilidad al parámetro que estima.
- ***Eficiente:*** Es el estimador que tiene la menor varianza entre todos los estimadores posibles.
- ***Suficiente:*** Cuando incluye toda la información que la muestra puede proporcionar acerca del parámetro.

2.2.2. Representación del modelo del dominio

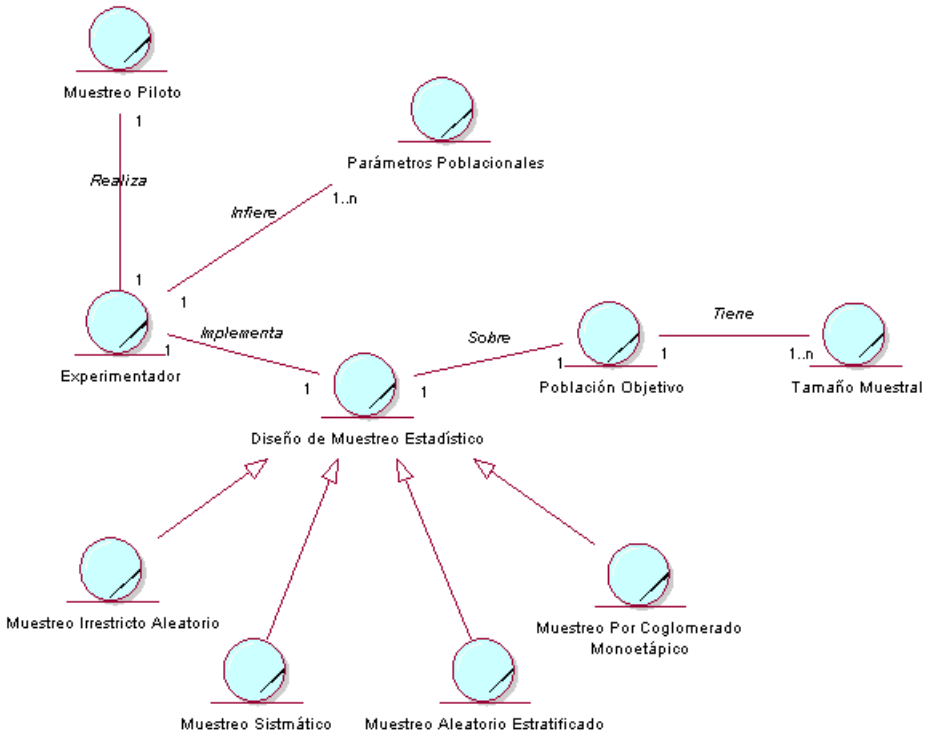


Figura 2.1: Diagrama de clases del modelo de objetos del dominio

2.2.3. Reglas del negocio a considerar

Las reglas del negocio constituyen un conjunto de políticas o patrones a ser cumplidos con el fin de garantizar el adecuado funcionamiento del proceso del negocio. En el ambiente de trabajo asociado al sistema se han determinado las siguientes reglas del negocio:

- La población bajo estudio será clasificada en dos tipos posibles: *Finita*, *Infinita*.
- El tipo de dato a analizar será clasificado en dos tipos posibles: *Cuantitativo*, *Cualitativo*.
- Un estudio de muestreo comprende el cálculo del tamaño muestral y la estimación del los parámetros poblaciones.
- Los parámetros poblaciones son: *Media Poblacional*, *Desviación Estándar*, *Total Poblacional*, *Varianza Muestral*, *Límite de error para la media*, *Límite de error para el total*, *Intervalo de confianza para la media*, *Intervalo de confianza para el total*.
- El estudio de muestreo deberá ser implementado para una sola variable de análisis.
- Los diseños de muestreos son: *Irrestricto Aleatorio (Muestreo Simple)*, *Aleatorio Estratificado*, *Sistemático*, *Por Conglomerados en una etapa*, *Por Conglomerados en dos etapas*.
- Cálculo del tamaño muestral se realizará en concordancia con el diseño de muestreo establecido por el experimentador.
- La determinación del valor de la muestra estará asociado al error máximo permisible.
- En cada reporte muestral se establecen las unidades seleccionadas en el cálculo del número de elementos de la muestra.

2.3. Descripción del sistema propuesto

2.3.1. Concepción general del sistema

El móvil principal de este trabajo de diploma es la concepción de un producto de software para el cálculo del tamaño de muestra en el proceso de investigación científico. El sistema propuesto

ofrece un ambiente agradable de fácil manejo con la finalidad de garantizar un adecuado proceder a partir del conocimiento mínimo de diferentes aspectos teóricos involucrados en el campo de acción de la estadística inferencial.

El sistema en su fase inicial dispone de tres opciones relacionadas con el análisis de datos a) Confeccionar un nuevo estudio de muestreo b) Confeccionar un nuevo estudio piloto c) Insertar los datos muestrales. Al seleccionar la ***opción a)*** el sistema permitirá la introducción de los datos correspondientes al estudio de muestreo, entre las cuales se encuentran:

- Título del estudio de muestreo.
- Nombre de la población bajo estudio
- Definición de la variable de análisis
- Tipificación del dato a analizar
- Clasificación de la población
- Total de unidades en la población

Posteriormente, el investigador podrá establecer el diseño de muestreo que, según su criterio, corresponda con la población objetivo, de igual forma se le muestran dos opciones que permiten definir el modo de cálculo del tamaño muestral.

Los diseños de que dispone son los siguientes:

- Muestreo Irrestricto Aleatorio
- Muestreo Aleatorio Estratificado
- Muestreo Sistemático
- Muestreo por Conglomerado en una etapa (Monoetápico)

De igual forma las opciones de determinación de la muestra son:

- Estimación de la media poblacional
- Estimación del total poblacional

Una vez introducidos los datos referentes al entorno general del estudio de muestreo en cuestión el investigador estará en condiciones de efectuar el cálculo del número de elementos de la muestra, para esto procede a la entrada de los datos, la cual, se realizará de forma específica en relación directa con el diseño de muestreo seleccionado anteriormente.

En el caso de seleccionar el muestreo aleatorio simple o sistemático le serán solicitados los valores de varianza o proporción según la tipificación del dato en análisis establecida previamente y error máximo permisible (estos valores son requeridos para todos los diseños). Si la elección fue muestreo aleatorio estratificado contará con tres opciones de asignación del tamaño de muestra *Óptima*, *Neyman o Tshuprow*, *Proporcional* y una opción relacionada con los costos de la encuesta donde se requiere del costo total de la encuesta y un costo fijo preestablecido, además se deberán ingresar los datos correspondientes por estratos (nombre del estrato, total de unidades que lo forman, varianza o proporción, costo de selección por unidad).

En el caso del muestreo por conglomerado en una etapa deberá ingresar el total de conglomerados que conforman la población, sin embargo si la elección corresponde al muestreo por conglomerados en dos etapas se solicitan además, los valores de la varianza entre las medias dentro de los conglomerados, varianza de los elementos dentro del conglomerado, costo asociado al seleccionar un conglomerado y costo asociado al seleccionar un elemento dentro del conglomerado.

Realizada la entrada de datos para el diseño de muestreo se procede al cálculo de la muestra el cual es visualizado mediante un reporte mostrándose los elementos seleccionados aleatoriamente que la conforman, así como, los valores introducidos por el investigador en el desarrollo del presente estudio.

Si se selecciona la ***opción b)*** se pondrá a disposición del usuario una ventana donde se solicitan los valores del error máximo permisible, el nivel de confianza y un listado de valores numéricos que, en dependencia con la tipificación del dato en análisis podrá ser de longitud 4 o 30 unidades. Seguidamente se estará en condiciones de determinar el número de elementos de la muestra piloto visualizado a través de un reporte que contendrá los elementos introducidos de forma preliminar conjuntamente con los valores de varianza muestral, longitud del intervalo, límites del intervalo de la media muestral (inferior y superior) además contará con los elementos seleccionados aleatoriamente que junto a los introducidos por el investigador constituyen el nuevo tamaño muestral.

La ***opción c)*** del análisis de los datos está relacionada con las alternativas explicadas con anterioridad, puesto que, constituye un segundo momento de utilización del software, donde, en correspondencia con el diseño implementado, el valor de la muestra obtenido y las unidades seleccionadas al azar, le son pedidos al investigador los valores recopilados una ves efectuado el trabajo de campo (aplicación del modelo de encuesta).

De manera independiente a la técnica de muestreo se debe introducir un listado de valores cumpliendo efectivamente con los obtenidos de la población encuestada. Para el diseño aleatorio estratificado se introducen dichos valores en relación al tamaño de la muestra asignado en cada subpoblación (estrato). En el caso del muestreo por conglomerado tanto en una como en dos etapas primero se establecerán los conglomerados partiendo de la selección aleatoria (número asignado por el diseño muestral), se incorporan el nombre y el total de unidades que lo forman o le fue asignado en el diseño muestral. Luego se podrán estimar los parámetros poblacionales visualizados mediante un reporte, dichos estadísticos de la población son los siguientes:

- Media poblacional
- Total poblacional
- Varianza muestral
- Intervalos de confianza para la media poblacional
- Intervalos de confianza para el total poblacional
- Límite de error en la estimación de la media poblacional
- Límite de error en la estimación del total poblacional
- Varianza de la estimación de la media poblacional
- Varianza de la estimación del total poblacional
- Desviación estándar de la media poblacional

2.3.2. Requerimientos Funcionales

Los requerimientos funcionales están determinados por el conjunto de especificaciones que definen completamente el comportamiento del sistema dígase atributos del sistema o del entorno del

sistema, entradas del sistema, salidas del sistema y sus funciones.

Los requerimientos funcionales del sistema propuesto son los siguientes:

1. Seleccionar opción de análisis de datos.
2. Título del estudio de muestreo
3. Nombre de la población bajo estudio
4. Definición de la variable de análisis
5. Tipificación del dato en estudio
6. Clasificación de la población
7. Total de unidades en la población
8. Establecer el diseño de muestreo
9. Establecer el modo de cálculo para el tamaño muestral
10. Introducir el error máximo permisible
11. Insertar el valor de la varianza poblacional estimada o el valor de la proporción poblacional estimada
12. Calcular tamaño el muestral para el Muestreo Irrestricto Aleatorio
13. Visualizar reporte el muestral para el Muestreo Irrestricto Aleatorio
14. Calcular tamaño muestral para el Muestreo Sistemático
15. Visualizar reporte muestral para el Muestreo Sistemático
16. Seleccionar el método de asignación de la muestra
17. Insertar el nombre del estrato
18. Insertar el total poblacional
19. Insertar la varianza o proporción poblacional
20. Insertar el costo de selección por unidad

21. Establecer costo total de la encuesta
22. Establecer costo fijo de la encuesta
23. Calcular tamaño muestral para el Muestreo Aleatorio Estratificado
24. Visualizar reporte muestral para el Muestreo Aleatorio Estratificado
25. Insertar el total de conglomerados
26. Calcular tamaño muestral para el Muestreo Por Conglomerado Monoetápico
27. Visualizar reporte muestral para el Muestreo Por Conglomerado Monoetápico
28. Título del estudio de muestreo piloto
29. Nombre de la población bajo estudio piloto
30. Definición de la variable de análisis para el estudio piloto
31. Insertar el tipo de dato de análisis
32. Insertar el nivel de confianza deseado
33. Insertar el nivel error máximo permisible
34. Calcular tamaño muestral para el Muestreo Piloto
35. Visualizar reporte muestral para el Muestreo Piloto
36. Insertar los datos muestrales
37. Calcular los parámetros poblacionales
38. Consultar la ayuda del sistema informático

2.3.3. Requerimientos No Funcionales

Los requerimientos no funcionales describen las restricciones del sistema o del proceso de desarrollo; no se refieren directamente a las funciones específicas que entrega el sistema, sino a las propiedades emergentes de éste como la fiabilidad, la respuesta en el tiempo y la capacidad

de almacenamiento. De forma alternativa, se definen las restricciones del sistema como la capacidad de los dispositivos de entrada/salida, en cuanto a prestaciones, atributos de calidad y la representación de datos que se utiliza en la interfaz del sistema.

Requerimientos de apariencia o interfaz

- La interfaz de la aplicación se diseñará de forma tal que garantice la uniformidad requerida para desempeñar adecuadamente el estudio de muestreo en cuestión por parte de investigador.
- Dado el usuario final que interactuará con el sistema se necesitará una apariencia apropiada que le mantenga orientado en todo momento lo cual le permita transitar con gran facilidad dentro del sistema de un punto a otro.
- Confección del sistema de reportes evitando sobrecargar el nivel de información proporcionada al investigador a través de un diseño sencillo.

Requerimientos de usabilidad

- La explotación del sistema agilizará la actividad científica del centro donde se ponga en práctica gracias a las comodidades brindadas por esta aplicación, la rapidez y organización presentado en sus cálculos, análisis y sistemas de reportes.
- El sistema contendrá un guía de usuario que comprenda los aspectos generales a tener en cuenta para el trabajo con la aplicación.
- El usuario que utilice la aplicación no necesitará de profundos conocimientos de la estadística para poder interactuar con el mismo de forma precisa.
- Se establecen diversas políticas que garantizan la entrada correcta de los datos para alcanzar la confiabilidad deseada de los resultados expuestos en la investigación.

Requerimientos de rendimiento

- Tiempos de respuesta del sistema adecuados al volumen de datos proporcionados por el usuario, estos serán generalmente instantáneos o de corta duración.
- Los cálculos a desempeñar por el sistema no requieren un alto nivel de procesamiento lo cual hace que la capacidad de procesamiento se comporte relativamente baja.

Requerimientos de soporte

- El sistema dispondrá de una arquitectura de diseño flexible que facilite su perfeccionamiento gradual, así como la incorporación de funcionalidades que aumenten las potencialidades del mismo.

Requerimientos de software

- Para llevar a cabo el proceso de instalación del sistema propuesto se debe disponer Windows 95 o superior y requiere de la instalación de Internet Explorer 6 o versiones superiores.

Requerimientos de hardware

- Para el proceso de instalación del sistema propuesto se necesitan como mínimo los siguientes requerimientos de hardware:
    - Procesador Pentium
    - 128 MB de RAM
    - 1 GB de HDD libre
    - Mouse

Restricciones del diseño e implementación

- Para garantizar una mejor documentación del sistema, se utiliza en el análisis, diseño e implementación del sistema propuesto el UML (Unified Modeling Language) y como herramienta de apoyo a este lenguaje de modelación se hace uso del Rational Rose.

2.3.4. Modelo de casos de uso del sistema

En el modelo de casos de uso del sistema quedan reflejados los diferentes casos de uso y su interrelación incluyendo en cada caso las descripciones correspondientes según el caso uso en cuestión.

Actores del sistema

Un actor describe un rol o comportamiento dentro del sistema a través de su interacción con el mismo.

Actor	Descripción
Experimenador	Usuario que interactúa con el sistema con la finalidad de llevar a cabo un estudio de muestreo estadístico, por lo tanto su accionar dentro sistema abarca todas funcionalidades previstas en el mismo. **Requisitos funcionales asociados:** R1, R2, R3,. . . , R38

Tabla 2.1: Descripción de los actores del sistema

Casos de uso del sistema

Cada forma en que los actores usan el sistema se representa con un caso de uso. Los casos de uso son ”fragmentos”de funcionalidad que el sistema ofrece para aportar un resultado de valor para sus actores (Jacobson, 2000).

La interacción existente entre los actores y sistema se describe a partir de los casos de uso, de esta forma el caso de uso refiere una secuencia de acciones a realizar por el sistema con el fin de satisfacer las necesidades de los actores.

Los casos de uso para el sistema propuesto son los siguientes:

1. Gestionar los datos del estudio del muestreo
2. Confeccionar el reporte muestral para el muestreo Irrestricto Aleatorio
3. Confeccionar el reporte muestral para el muestreo Sistemático
4. Gestionar los datos para el diseño muestral Aleatorio Estratificado
5. Confeccionar el reporte muestral para el muestreo Aleatorio Estratificado
6. Confeccionar el reporte muestral para el muestreo por Conglomerados Monoetápico
7. Gestionar los datos del diseño muestral Piloto
8. Confeccionar el reporte muestral del estudio Piloto
9. Insertar los datos generales para cada diseño muestral
10. Insertar los datos muestrales
11. Consultar la ayuda del sistema

2.3.5. Diagrama de casos de uso del sistema

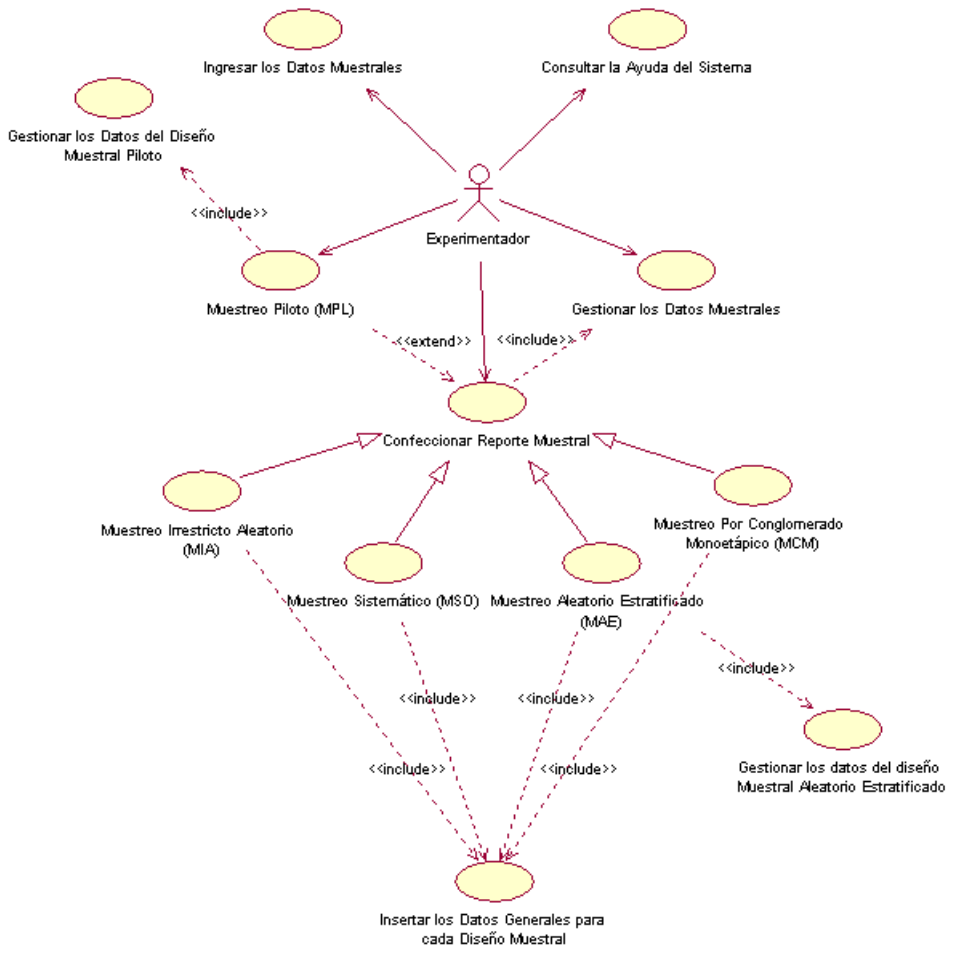


Figura 2.2: Diagrama de casos de uso del sistema

2.3.6. Descripción de los casos de uso del sistema

Caso de uso	Gestionar los datos muestrales
Actores	Investigador
Propósito	Posibilitar el cálculo del tamaño muestral
Resumen El caso de uso se inicia cuando el usuario accede al sistema y selecciona la opción del análisis de los datos confeccionar un nuevo estudio de muestreo. El sistema solicita el título de la investigación, el nombre de la población bajo estudio, definición de la variable de análisis, tipificación del dato en estudio, clasificación de la población, el total de unidades en la población. A continuación, el sistema solicita al usuario las opciones de diseño muestral, además, las alternativas relacionadas con el modo de cálculo para el número de elementos de la muestra. El caso de uso finaliza cuando los datos han sido incorporados al sistema satisfactoriamente.	
Referencias	R1, R2, R3, R4, R5, R6, R7, R8, R9
Precondiciones	
Post-condiciones	-
Prototipo	Ver Anexo 2 Figura 3.1, Figura 3.2

Tabla 2.2: Caso de Uso: Gestionar los datos muestrales

Caso de uso	Confeccionar el reporte muestral para el Muestreo Irrestricto Aleatorio
Actores	Investigador
Propósito	Mostrar al experimentador los resultados del tamaño muestral y la tablas con los elementos seleccionados aleatoriamente de la población.
Resumen El caso de uso se inicia cuando el investigador incorpora al sistema los valores correspondientes al diseño de muestreo tales como son la varianza o la proporción poblacional y el error máximo permisible. El sistema para la realización de este caso de uso debe necesariamente ejecutar los casos de caso de uso 1 y 2. El caso de uso termina cuando el sistema realiza el cálculo del número de elementos de la muestra satisfactoriamente visualizando este resultado a través de un reporte.	
Referencias	R12, R13, CU1 (include), CU2 (include)
Precondiciones	
Post-condiciones	-
Prototipo	Ver Anexo 2 Figura 3.3,Figura 3.4

Tabla 2.3: Caso de Uso: Confeccionar el reporte muestral para el Muestreo Irrestricto

Caso de uso	Confeccionar el reporte muestral para el Muestreo Sistemático
Actores	Investigador
Propósito	Mostrar al experimentador los resultados del tamaño muestral y la tablas con los elementos seleccionados aleatoriamente de la población.
Resumen El caso de uso se inicia cuando el investigador incorpora al sistema los datos correspondientes al diseño de muestreo tales como son la varianza o la proporción poblacional y el error máximo permisible. El sistema para la realización de este caso de uso debe necesariamente ejecutar los casos de caso de uso 1 y 2. El caso de uso termina cuando el sistema realiza el cálculo del número de elementos de la muestra satisfactoriamente visualizando este resultado a través de un reporte.	
Referencias	R14, R15, CU1 (include), CU2 (include)
Precondiciones	
Post-condiciones	-
Prototipo	Ver Anexo Figura 3.5, Figura 3.6

Tabla 2.4: Caso de Uso: Confeccionar el reporte muestral para el Muestreo Sistemático

Caso de uso	Gestionar los datos del diseño muestral Aleatorio Estratificado
Actores	Investigador
Propósito	Permitir la determinación del tamaño de muestra
Resumen El caso de uso se inicia cuando el usuario se dispone a incorporar al sistema los datos correspondientes al diseño muestral aleatorio estratificado. El sistema le solicita los datos requeridos para el cálculo del número de elementos de la muestra, es decir, la selección del método de asignación de la muestra, insertar el nombre del estrato, valor del total poblacional, la varianza o la proporción poblacional, costo de seleccionar una unidad y en caso de ser requerido establecer el costo total de la encuesta junto a un costo prefijado. El caso de uso finaliza cuando los datos solicitados han sido introducidos correctamente en el sistema.	
Referencias	R15, R16, R17, R18, R19, R20
Precondiciones	
Post-condiciones	-
Prototipo	Ver Anexo Figura 3.7

Tabla 2.5: Caso de Uso: Gestionar los datos del diseño muestral Aleatorio Estratificado

Caso de uso	Confeccionar el reporte muestral para el diseño Aleatorio Estratificado
Actores	Investigador
Propósito	Mostrar al experimentador los resultados del tamaño muestral y la tablas con los elementos seleccionados aleatoriamente de la población y la asignación por estratos.
Resumen El caso de uso se inicia cuando el investigador incorpora al sistema los datos correspondientes al diseño de muestreo tales como son la selección del método de asignación de la muestra, insertar el nombre del estrato, valor del total poblacional, la varianza o la proporción poblacional, costo de seleccionar una unidad y en caso de ser necesario establecer el costo total de la encuesta junto a un costo prefijado. El sistema para la realización de este caso de uso debe necesariamente ejecutar los casos de uso 2 y 3 respectivamente. El caso de uso termina cuando el sistema realiza el cálculo del número de elementos de la muestra satisfactoriamente el cual es visualizado a través de un reporte.	
Referencias	R23, R24, CU3 (include)
Precondiciones	
Post-condiciones	-
Prototipo	Ver Anexo 2 Figura 3.8

Tabla 2.6: Caso de Uso: Confeccionar el reporte muestral para el diseño Aleatorio Estratificado

Caso de uso	Confeccionar el reporte muestral para el diseño Por Conglomerados Monoetápico
Actores	Investigador
Propósito	Mostrar al experimentador los resultados del tamaño muestral y la tabla con los elementos (conglomerados) seleccionados aleatoriamente de la población.
Resumen El caso de uso se inicia cuando el investigador incorpora al sistema los datos correspondientes al diseño de muestreo tales como son el total de conglomerados que conforman la población, la varianza o proporción poblacional, el error máximo permisible. El sistema para la realización de este caso de uso debe necesariamente ejecutar los casos de uso 1, 2 respectivamente. El caso de uso termina cuando el sistema realiza el cálculo del número de elementos de la muestra satisfactoriamente, visualizando este resultado mediante un reporte.	
Referencias	R25, R26, R27, CU1 (include), CU2 (include)
Precondiciones	
Post-condiciones	-
Prototipo	Ver Anexo 2 Figura 3.9, Figura 3.10

Tabla 2.7: Caso de Uso: Confeccionar el reporte muestral para el diseño Por Conglomerados Monoetápico

Caso de uso	Gestionar los datos para el diseño muestral Piloto
Actores	Investigador
Propósito	Permitir la determinación del tamaño muestral
Resumen El caso de uso se inicia cuando el usuario se dispone a incorporar al sistema los datos correspondientes al diseño muestral Piloto. El sistema solicita los datos correspondientes con el título del estudio piloto, la localidad donde se realiza el estudio, la tipificación de dato de análisis, nivel de confianza deseado por el experimentador junto al error máximo permisible y el total poblacional. El caso de uso finaliza cuando los datos han sido introducidos correctamente en el sistema.	
Referencias	R28, R29, R30, R31, R32, R33
Precondiciones	
Post-condiciones	-
Prototipo	Ver Anexo 2 Figura 3.11, Figura 3.12

Tabla 2.8: Caso de Uso: Gestionar los datos para el diseño muestral Piloto

Caso de uso	Confeccionar el reporte muestral para el Muestreo Piloto
Actores	Investigador
Propósito	Mostrar al experimentador los resultados del tamaño muestral piloto y la tabla con los elementos seleccionados aleatoriamente de la población.
Resumen El caso de uso se inicia cuando el usuario accede al sistema y selecciona la opción del análisis de los datos confeccionar nuevo estudio piloto seguidamente el sistema solicita los datos correspondientes al diseño de muestreo tales como son el nivel de confiabilidad, el error máximo permisible, además requiere de un listado de valores numéricos. El sistema para la realización de este caso de uso requiere necesariamente de la ejecución del caso de uso 6. El caso de uso termina cuando el sistema realiza el cálculo del número de elementos de la muestra de forma tal que garantice determinados especificaciones del modelo matemático, este resultado es visualizado mediante un reporte.	
Referencias	R34, R35, CU6 (include)
Precondiciones	
Post-condiciones	-
Prototipo	Ver Anexo 2 Figura 3.13

Tabla 2.9: Caso de Uso: Confeccionar el reporte muestral para el Muestreo Piloto

Caso de uso	Insertar los datos generales para cada diseño muestral
Actores	Investigador
Propósito	Incorporar al sistema los datos necesarios para el cálculo del tamaño muestral
Resumen El caso de uso inicia cuando el usuario se dispone a incorporar los datos para cualquier diseño muestral previsto en la aplicación. El sistema, en correspondencia al diseño muestral, solicita un juego de datos específico, de manera simultánea le serán pedidos al usuario un conjunto de valores que resultarán comunes para cada técnica de muestreo que se pretenda implementar sobre la población objeto de estudio. Los valores solicitados por el sistema son los siguientes, la varianza poblacional, la proporción poblacional y el error máximo permisible. El caso de uso finaliza cuando los datos solicitados han sido introducidos correctamente en el sistema.	
Referencias	R10, R11
Precondiciones	
Post-condiciones	-
Prototipo	Ver la Figura 3.5, Figura 3.7, Figura 3.9, Figura 3.11

Tabla 2.10: Caso de Uso: Insertar los datos generales para cada diseño muestral

Caso de uso	Insertar los datos Muestrales
Actores	Investigador
Propósito	Permitir al investigador la estimación de los parámetros poblacionales
Resumen El caso de uso se inicia cuando el usuario selecciona la opción de análisis de datos correspondiente, acto seguido el sistema le solicita el fichero de datos requerido mediante una ventana de diálogo. En dependencia del fichero de datos seleccionado por el usuario el sistema brindará la interfaz gráfica que le permita introducir los datos según el diseño muestral. El caso de uso finaliza una vez que el usuario introduzca los valores en el sistema de forma satisfactoria.	
Referencias	R37
Precondiciones	
Post-condiciones	-
Prototipo	Ver Anexo 2 Figura 3.14, Figura 3.15, Figura 3.16

Tabla 2.11: Caso de Uso: Insertar los datos Muestrales

Caso de uso	Consultar la ayuda del sistema
Actores	Investigador
Propósito	Proporcionar información de utilidad al experimentador respecto al uso y funcionamiento del sistema
Resumen El caso de uso se inicia cuando el usuario solicita la ayuda de algún tópico relacionado con el modo de uso del software o el desconocimiento de ciertas cuestiones ligadas a la estadística. En dependencia con la necesidad del usuario el sistema brindará la información que se ajuste adecuadamente a sus necesidades. El caso de uso finaliza una vez que el usuario realice la consulta de la temática deseada de forma satisfactoria.	
Referencias	R38
Precondiciones	
Post-condiciones	-
Prototipo	Ver Anexo 2 Figura 3.17

Tabla 2.12: Caso de Uso: Consultar la ayuda del sistema

2.4. Construcción de la solución propuesta

2.4.1. Diagrama de clases del diseño

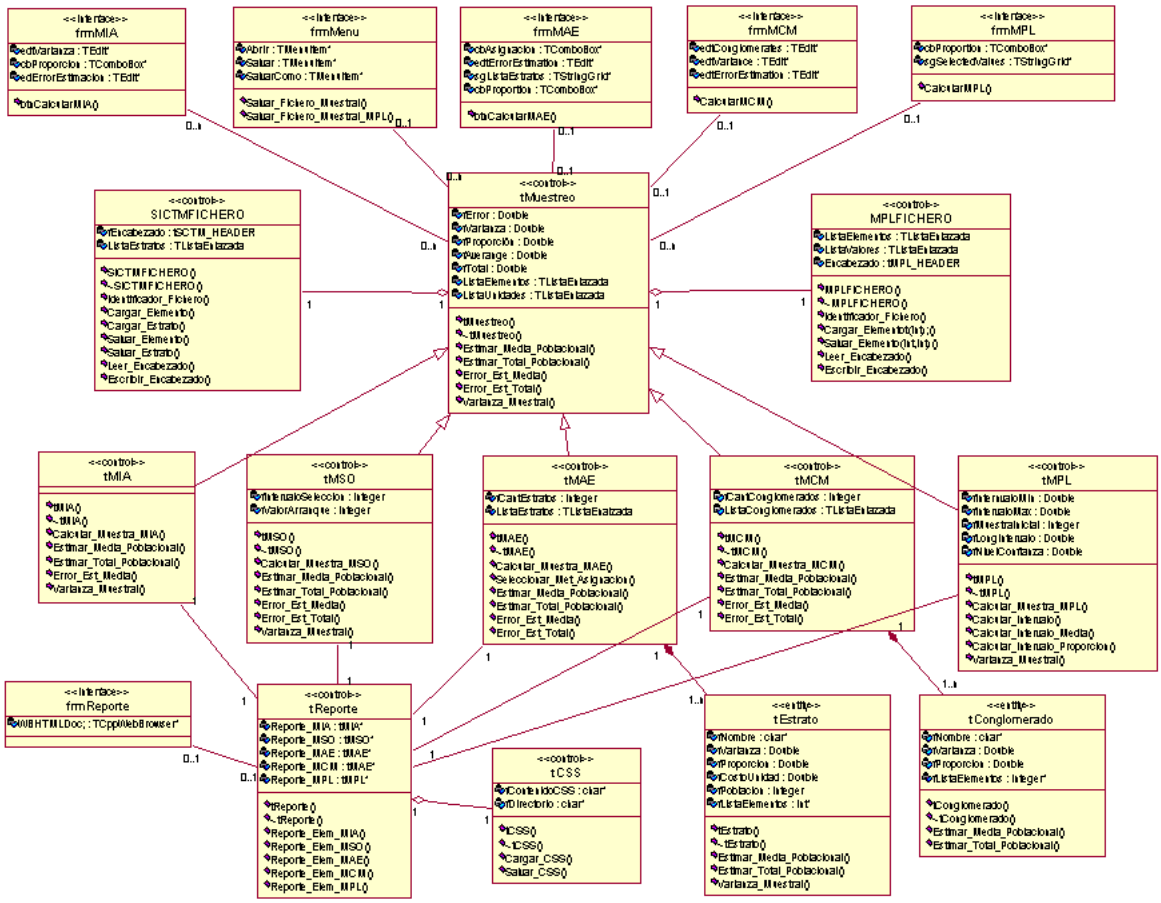


Figura 2.3: Diagrama de clases del diseño

2.4.2. Diagrama de archivo

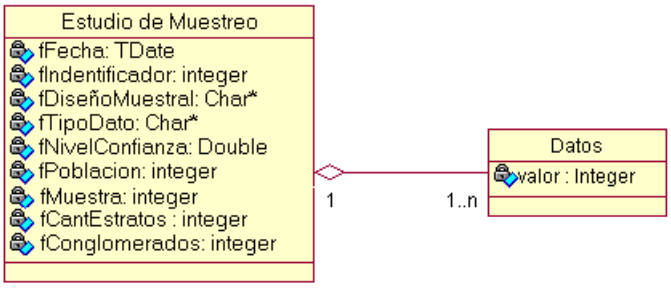


Figura 2.4: Diagrama de clases persistente. Estudio de Muestreo

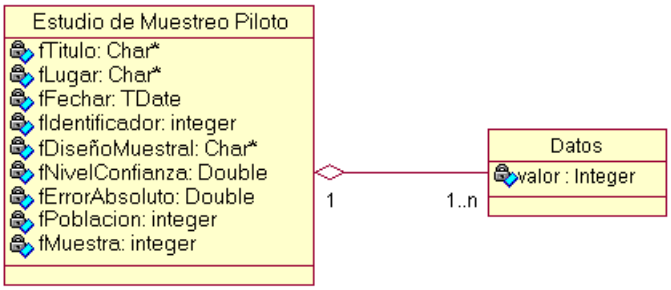


Figura 2.5: Diagrama de clases persistente. Estudio de Muestreo Piloto

2.4.3. Estructura de los archivos

ARCHIVO ESTUDIO DE MUESTREO

Encabezado	Datos

Tabla 2.13: Estructura del Archivo Estudio de Muestreo

ENCABEZADO DEL ARCHIVO ESTUDIO DE MUESTREO

fFecha: TDate
fIdentificador:Integer
fDiseñoMuestral: Char*
fTipoDato: Char*
fNivelConfianza: Double
fPoblación: Integer
fMuestra: Integer
fCantEstratos: Integer
fCantConglomerados: Integer

DESCRIPCIÓN DE LOS CAMPOS DEL ENCABEZADO:

- **Fecha**: Día, Mes y Año en que fue iniciado el estudio muestral .
- **Identificador**: Especifica que el archivo fue creado por el sistema.
- **Diseño Muestral**: Diseño de muestreo establecido por el experimentador.
- **Tipo de Dato de Análisis**: Especifica el tipo de dato objeto de estudio en la investigación.
- **Nivel de Confianza**: Nivel de confianza introducido por el experimentador para realizar las estimaciones muestrales.
- **Población**: Total de unidades comprendidas en el área de investigación.
- **Muestra**: Número total de elementos seleccionados aleatoriamente.
- **Cantidad de Estratos**: Número total de estratos que conforman la población.
- **Cantidad de Conglomerados**: Número total de conglomerados que conforman la población.

ESTRUCTURA DEL REGISTRO DATO

fNombreEstrato: Char*
fCostoUnidad: Double
fProporcion: Double
fDesviacionEstandar: Double
fVarianzaMuestral: Double
fPoblacion: Integer
fMuestra: Integer

DESCRIPCIÓN DE LOS CAMPOS DEL REGISTRO DATO:

- **Nombre del Estrato**: Nombre correspondiente al estrato i-ésimo.
- **Costo por Unidad**: Costo asociado a la selección de un elemento dentro del estrato i-ésimo.
- **Proporción**: Proporción de la característica en estudio dentro del estrato i-ésimo.
- **Desviación Estándar**: Valor de desviación estándar correspondiente al estrato i-ésimo.
- **Varianza Muestral**: Valor calculado entre las unidades dentro del estrato i-ésimo.
- **Población**: Número total de unidades dentro del estrato i-ésimo.
- **Muestra**: Número de elementos seleccionados aleatoriamente dentro del estrato i-ésimo.

ESTRUCTURA DEL REGISTRO DATO

fElemento: Integer

DESCRIPCIÓN DEL CAMPO DEL REGISTRO DATO:

- **Elemento**: Elemento seleccionado dentro del marco muestral determinado por el tamaño de muestra.

ENCABEZADO DEL ARCHIVO ESTUDIO DE MUESTREO PILOTO

fTitulo: Char*
fLugar: Char*
fFecha: TDate
fIdentificador:Integer
fDiseñoMuestral: Char*
fTipoDato: Char*
fNivelConfianza: Double
fErrorAbsoluto: Double
fPoblación: Integer
fMuestra: Integer

DESCRIPCIÓN DE LOS CAMPOS DEL ENCABEZADO:

- **Título**: Título asociado a la investigación.
- **Lugar**: Área geográfica donde se enmarca la investigación.
- **Fecha**: Día, Mes y Año en que fue iniciado el estudio muestral .
- **Identificador**: Especifica que el archivo fue creado por el sistema.
- **Diseño Muestral**: Diseño de muestreo establecido por el experimentador.
- **Tipo de Dato de Análisis**: Especifica el tipo de dato objeto de estudio en la investigación.
- **Nivel de Confianza**: Nivel de confianza introducido por el experimentador para realizar las estimaciones muestrales.
- **Error Absoluto**: Valor del error máximo permisible que introduce el experimentador en sus estimaciones.
- **Población**: Total de unidades comprendidas en el área de investigación.
- **Muestra**: Número total de elementos seleccionados aleatoriamente.

2.4.4. Diagrama de implementación

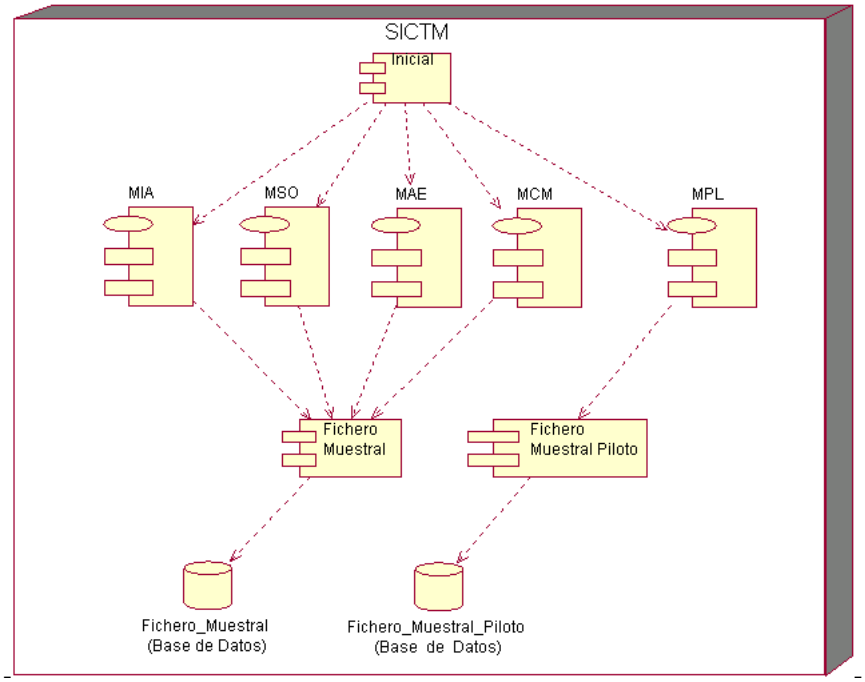


Figura 2.6: Diagrama de implementación

2.5. Principios de diseño

2.5.1. Estándares en la interfaz de la aplicación

La calidad de la interfaz de usuario constituye uno de elementos con más relevancia en la aceptación y puesta en uso de un software.

El sistema propuesto está dirigido a cualquier usuario que aborde el cálculo del tamaño de muestras, por esta razón, se presenta una interfaz sencilla de manera que se garantice un proceder adecuado en la interacción con el mismo.

Teniéndose en cuenta el estudio de diferentes software vinculados con el ámbito investigativo y después de tomar a consideración la experiencia empírica adquirida por desarrolladores de software desktop, se determinó que la interfaz requerida debe caracterizarse por la no utilización de imágenes sobrecargadas, así como, evitar el uso sobredimensionado del color y de igual forma se debe minimizar el empleo de animaciones que proporcionen un elemento distractor al usuario.

La tipografía utilizada es MS Sans Serif , de estilo normal, color negro y tamaño 12, además, el desarrollo de las interfaces visuales están basadas en el estándar de ventanas de Windows.

2.5.2. Tratamiento de errores

Durante el proceso de diseño e implementación del sistema se ha tenido en cuenta las validaciones correspondientes para evitar la introducción de información errónea por parte del usuario. Cuando el usuario teclea los datos siempre se valida esa información de modo tal que cuando se requiera de un dato tipo numérico no se cometa el error de entrar uno tipo texto y viceversa, además los botones asociados al formulario no se activan hasta que todas las cajas de edición sean llenadas.

Cada uno de los procedimientos de muestreo comprendidos en la aplicación necesita de cierta cantidad de valores para determinar el tamaño muestral y posteriormente visualizar los reportes con la información correspondiente al diseño muestral, por lo tanto el sistema ante la entrada incorrecta de alguno de estos datos emite mensajes de error con la información adecuada y ajustada al nivel de conocimiento estadístico del usuario.
Cuando el usuario selecciona la opción de insertar los datos muestrales, el sistema a partir del archivo de datos seleccionado muestra un ambiente visual que brinda facilidades para la inserción de los valores, teniendo en cuenta la validación de que los mismos correspondan tan solo al tipo de dato numérico.

2.5.3. Concepción general de la ayuda

La ayuda constituye una parte importante del sistema pues le proporciona al usuario toda la información necesaria respecto al uso del sistema y los conceptos de la estadística que deben ser conocidos por el mismo.

Con este fin, se concibe una ayuda organizada por módulos según los diseños de muestreo previstos en el sistema. Para cada tópico de la ayuda se explica brevemente los conceptos vinculados a cada diseño muestral, además de disponer de un ejemplo práctico que ilustre al usuario en su interacción con el sistema. Desplegando cada módulo aparecen todas las opciones que corresponden a cada tópico y en el panel a la derecha, su explicación.

Es importante destacar que durante el desarrollo de un estudio de muestreo o la inserción de datos muestrales el sistema brinda opciones de ayuda que referencian específicamente el módulo de trabajo que se encuentre activo. Además, el sistema de ayuda brinda posibilidades de búsqueda de palabras o términos estadísticos que sean solicitados por el usuario.

2.6. Conclusiones Parciales

En el desarrollo del presente capítulo se ha enfatizado en la utilización de la metodología RUP, lo cual a permitido desglosar detalladamente el sistema propuesto a partir de la descripción de sus funcionalidades en la captura de requisitos tanto funcionales como no funcionales.

Como otro resultado de relevancia se puede mencionar la definición del diagrama de clases del modelo de objetos del dominio y el diagrama de casos de uso del sistema. La descripción de los casos de uso como un punto culminante del desarrollo de este capítulo a posibilitado una mejor comprensión acerca del entorno en que se desenvuelve el sistema y la forma en que interactúa con usuario.

Se mostraron los resultados de la etapa de diseño del sistema. Se presentó el diagrama de clases del diseño, los diagramas del modelo lógico y físico de datos para una mayor comprensión del funcionamiento de la base de datos, el diagrama de implementación para representar los elementos fundamentales de la misma, así como los principios de diseño utilizados

Capítulo 3

Análisis de factibilidad y validación de la solución propuesta

3.1. Introducción

Este capítulo trata de diferentes aspectos relacionados al estudio de la factibilidad del producto. Se estiman el esfuerzo humano y el tiempo de desarrollo que se requieren para la elaboración del mismo, así como los costos y los beneficios tangibles e intangibles que reporta la utilización del sistema. Se realiza al análisis entre los costos y los beneficios para concluir si es o no factible el desarrollo del sistema, para ello se emplea el método de puntos de función del modelo de COCOMO II.

Se lleva a cabo la validación de la solución propuesta basándose en la opinión de los especialistas, mediante la aplicación de encuestas personalizadas, a partir del establecimiento de tres grupos de expertos *(Contenido, Instruccional, Informática)*. Finalmente, se efectúa el procesamiento de la información con el ayuda del software SPSS v15. Los métodos previstos para fundamentar la validación del sistema fueron el coeficiente Alpha de Cronbach y la Prueba no Paramétrica W. de Kendall.

3.2. Planificación por puntos de función

Para estimar el tamaño del producto y el esfuerzo asociado a su desarrollo se utiliza el método de estimación por puntos de función.

Para realizar el cálculo de los costos de desarrollo del sistema se analizan las cantidades de entradas, salidas, peticiones, archivos lógicos e interfaces externas preliminares que tiene el sistema. También se tuvo en cuenta la conversión al C++, que es de 53 puntos.

Después de efectuado el estudio se obtienen los siguientes resultados:

3.3. Determinación de los costos

Cálculo del esfuerzo, tiempo de desarrollo, cantidad de hombres y costo.

Multiplicador de esfuerzos

$$\begin{aligned} EM &= \prod_{j=1} EM_i = RCPX * RUSE * PDIF * PERS * PREX * FCIL * SCED \\ EM &= \prod_{j=1} EM_i = 1 * 1 * 1 * 0{,}83 * 1 * 0{,}87 * 1 = 0{,}7221 \approx 0{,}72 \end{aligned}$$

Factores de escala

$$\begin{aligned} SF &= \sum SF_i = PREC + FLEX + TEAM + PMAT \\ SF &= \sum SF_i = 2,48 + 3,04 + 4,24 + 3,29 * 4,68 = 17,73 \end{aligned}$$

Valores de los coeficientes

$$\begin{aligned} A &= 2,94 \\ B &= 0,91 \\ C &= 3,67 \\ D &= 0,24 \\ E &= B + 0,01 * SF = 0,91 + 0,01 * 17,73 = 1,08 \\ F &= D + 0,2 * (E - B) = 0,24 + 0,2 * (1,08 - 0,91) = 0,275 \end{aligned}$$

Esfuerzo

$$PM = A * (MF)^E * EM = 2,94 * (8,538)^{1,08} * 0,72 = 21,45$$

Nombre de la entrada externa	Cantidad de ficheros	Cantidad de elementos de datos	Clasificación (Alto, Medio, Bajo)
Seleccionar opción de análisis de datos	1	9	Bajo
Insertar título del estudio de muestreo	1	9	Bajo
Insertar nombre de la población bajo estudio	1	9	Bajo
Definir la variable de análisis	1	9	Bajo
Tipificar el dato en estudio	1	9	Bajo
Clasificar la población	1	9	Bajo
Insertar total de unidades en la población	1	9	Bajo
Establecer el diseño de muestreo	1	9	Bajo
Establecer el modo de cálculo para el tamaño muestral	1	9	Bajo
Introducir el error máximo permisible	1	2	Bajo
Insertar el valor de la varianza poblacional estimada o el valor de la proporción poblacional estimada	1	2	Bajo
Establecer costo fijo de la encuesta	1	7	Bajo
Insertar título del estudio de muestreo piloto	1	6	Bajo
Insertar nombre de la población bajo estudio piloto	1	6	Bajo
Definir la variable de análisis para el estudio piloto	1	6	Bajo
Insertar el tipo de dato de análisis	1	6	Bajo
Insertar el nivel de confianza deseado	1	6	Bajo
Insertar el nivel error máximo permisible	1	6	Bajo

Tabla 3.1: Planificación: Entradas externas

Nombre de la entrada externa	Cantidad de ficheros	Cantidad de elementos de datos	Clasificación (Alto, Medio, Bajo)
Calcular tamaño el muestral para el Muestreo Irrestricto Aleatorio	1	9	Bajo
Calcular tamaño muestral para el Muestreo Sistemático	1	9	Bajo
Calcular tamaño muestral para el Muestreo Aleatorio Estratificado	1	9	Bajo
Calcular tamaño muestral para el Muestreo Por Conglomerado Monoetápico	1	9	Bajo
Calcular tamaño muestral para el Muestreo Piloto	1	9	Bajo
Calcular tamaño muestral para el Muestreo Por Conglomerado Monoetápico	1	9	Bajo
Calcular tamaño muestral para el Muestreo Piloto	1	11	Bajo

Tabla 3.2: Planificación: Peticiones

Nombre del fichero interno	Cantidad de ficheros	Cantidad de elementos de datos	Clasificación (Alto, Medio, Bajo)
Muestreo Irrestricto Aleatorio	1	11	Bajo
Muestreo Sistemático	1	11	Bajo
Muestreo Aleatorio Estratificado	1	18	Bajo
Muestreo Por Conglomerado Monoetápico	1	12	Bajo
Muestreo Piloto	1	6	Bajo
Calcular tamaño muestral para el Muestreo Por Conglomerado Monoetápico	1	9	Bajo
Calcular tamaño muestral para el Muestreo Piloto	1	11	Bajo

Tabla 3.3: Planificación: Ficheros internos

Elementos	Bajos	X Peso	Medios	X Peso	Altos	X Peso	Subtotal de puntos de función
Ficheros lógicos internos	5	7	0	10	0	15	35
Ficheros de interfaces externas	0	5	0	7	0	10	0
Entradas externas	24	3	0	4	0	6	72
Salidas externas	5	4	0	5	0	7	20
Peticiones	5	3	0	4	0	6	15
Total							142

Tabla 3.4: Planificación: Punto de función

Características	Valor
Puntos de función desajustados	142
Lenguaje	C++
Instrucciones fuentes por puntos de función	53
Por ciento de la aplicación en cuanto a requerimientos funcionales	100 %
Total de Instrucciones fuentes	7526

Tabla 3.5: Planificación: Miles de instrucciones fuentes

Cálculo de:	Valor	Justificación
RCPX	1 (Nominal)	No se requiere de amplia documentación. El sistema tiene una moderada complejidad.
RUSE	1 (Nominal)	Se implementa código reusable para el aprovechamiento de este en toda la aplicación.
PDIF	1 (Nominal)	No tiene grandes restricciones en cuanto al tiempo de ejecución ya que el sistema podrá estar trabajando varias horas. El sistema no tiene limitación de memoria impuesta. La plataforma de aplicación tiene gran estabilidad.
PERS	0,83 (Alto)	Hay poco movimiento del personal.
PREX	1(Alto)	El equipo tiene buen dominio y posee conocimiento del lenguaje de programación. Con una experiencia de aproximadamente cuatro años.
FCIL	0,87 (Alto)	Se utilizan herramientas como: Borland Builder C++, así como CASE Rational Rose para la documentación, empleando como notación UML.
SCED	1 (Nominal)	La planificación se hace con moderada frecuencia.
PREC	2,48 (Alto)	El equipo de desarrollo posee una comprensión considerable de los objetivos del producto.
FLEX	3,04 (Nominal)	El sistema cuenta con alguna flexibilidad en relación con las especificaciones de los requerimientos preestablecidos y a las especificaciones de interfaz externa.
TEAM	3,29 (Nominal)	El equipo que va a desarrollar el sistema es altamente cooperativo.
RESL	4,24 (Nominal)	Teniendo en cuenta la alta experiencia que existe en el país acerca de este tipo de estudios existen algunos factores de riesgo.
PMAT	4,68 (Nominal)	Se encuentra en su primera etapa un poco avanzada.

Tabla 3.6: Costos: Factores de escalas

Cálculo del tiempo de desarrollo

$$TDEV = C * PM^F = 3,67 * (21,45)^{0,275} = 8,5$$

Cálculo de la cantidad de hombres

$$CH = PM/TDEV = 21,45/8,5 = 2,5$$

Recalculando

$$
\begin{aligned}
CH &= PM/TDEV \\
TDEV &= PM/CH \\
CH &= 1 \\
TDEV &= 21{,}45/1 = 21,45 \approx 22 \ \ meses
\end{aligned}
$$

Costo Se asume como salario promedio mensual 225 $

$$
\begin{aligned}
CHM &= 1 * SalarioPromedio = 1 * 225 = 225\$ \ \ /mes \\
Costo &= CHM * PM = 225 * 21{,}45 = 4826
\end{aligned}
$$

Los costos en los que se incurriría de desarrollarse el sistema serían:

Cálculo de:	Valor
Esfuerzo(PM)	21,45
Tiempo de desarrollo	22 meses
Cantidad de hombres	1
Costo	4826
Salario medio	225
RCPX	1
RUSE	1
PDIF	1
PREX	1
FCIL	0,87
SCED	1

Tabla 3.7: Costos totales

3.4. Beneficios tangibles e intangibles

Los proyectos informáticos, al igual que otros proyectos, generan efectos económicos, estos, según al tipo que correspondan presentan diversas clasificaciones, estas son: directos, indirectos, externos (***beneficios tangibles***) y ***beneficios intangibles***. Entiéndase como beneficios tangibles aquellos efectos que se alcanzan como consecuencia directa de la implantación del proyecto de manera que resulten perceptibles económicamente para la entidad generadora del mismo. Los beneficios intangibles son aquellos efectos apreciables por la comunidad como perjuicio o beneficio, pero al momento de ponderar en unidades monetarias esto resulta difícil o prácticamente imposible. Estos efectos se denominan también *efectos cualitativos*.
Los beneficios tangibles que pueden producir los proyectos informáticos son:

- Ahorro de horas-hombre (por no tener que contratar personal adicional o por el aumento de la productividad)
- Venta de información (venta de software)
- Ahorro en arriendo de oficinas (dado el paso a medios magnéticos de archivo a carpetas que existen físicamente)

- Ahorro en costos de operación (dejar de pagar servicios a empresas, disminución de costos de mantenimiento)
- Valor residual de los equipos
- Ahorro de horas-hombre del personal que actualmente labora en el sistema (aumento de la productividad)
- Ahorro de horas-hombre de los clientes (beneficio social)

Como beneficios intangibles asociados a la realización de un proyecto informático se pueden mencionar los siguientes:

- Mejora en la calidad de la información por la integridad, oportunidad de la información y la confiabilidad
- Mayor comodidad de los usuarios
- Mejor imagen de la institución
- Mejoramiento de las condiciones de trabajo del personal

3.5. Análisis de costos y beneficios

Este sistema, como resultado del presente trabajo de diploma, no implica costo alguno para la empresa, centro de estudio o cualquier entidad donde se pretenda implantar el software, sin embargo, al desarrollo de todo producto informático va asociado un costo y su justificación económica viene dado por los beneficios tangibles e intangibles que este produce.
La utilización de este nuevo sistema permite a los investigadores o experimentadores que requieran la determinación de tamaños de muestras desempeñar esta labor sobre un entorno visual agradable y que disponde de características propias para satisfacer las necesidades de los mismos relacionadas con el manejo de las tablas de números aleatorios, la confección de reportes y la recolección de datos. Además, posibilita aprovechar las potencialidades informáticas existentes en el centro, en función del mejoramiento del proceso investigativo, mediante la utilización de los medios computacionales. Para la realización de este sistema no fue necesaria una inversión en los medios técnicos.

3.6. Validación de la propuesta de solución

3.6.1. Procesamiento estadístico y análisis de los resultados

Para el procesamiento de la información se utilizó el paquete estadístico SPSS V.15 comenzando con un análisis descriptivo de la información recopilada con ayuda de los cuestionarios que fueron diseñados cumpliendo los requisitos de presentación, motivación, longitud adecuada, preguntas claras y simples, secuencia lógica, evitándose el empleo de fraseología negativa. En el Anexo 3 se adjuntan ejemplares de las encuestas aplicadas.

La forma de aplicación personal de la encuesta garantiza que la cantidad de encuestas entregadas o respondidas sean recuperadas referente a la proporción de no respondientes.
Para determinar la fiabilidad del cuestionario se utilizó el coeficiente Alpha de Cronbach, según las posibilidades del SPSS obteniéndose valores superiores a 0,799 lo que permite considerar que las puntuaciones percibidas en los diferentes items de los cuestionarios, están altamente interrelacionadas.

3.6.2. Valoraciones de acuerdo al criterio de expertos

Las opiniones de los expertos valoradas mediante métodos estadísticos, permite aumentar el nivel científico del trabajo realizado. En este sentido la evaluación se realiza a través de un sistema de procedimientos organizados y lógicos, dirigidos a obtener la información procedente de los expertos y su posterior análisis con el objetivo de tomar decisiones confiables.
En su generalidad, los métodos de evaluación de software requieren que los expertos revisen una amplia variedad de aspectos de los mismos. Según se refiere en la bibliografía consultada, en ocasiones, el número de criterios que los expertos deben evaluar es muy elevado y se pierde información debido al desconocimiento de determinados criterios que no son de dominio por parte del equipo que elabora los instrumentos para la valoración o porque el propio experto no tiene elementos que contribuyan a orientar su valoración. Sin embargo, muchos de ellos podían ser agrupados en dimensiones que apuntan a áreas del conocimiento más específicas, lo que redundaría en una mayor objetividad de los criterios de valor de los expertos.
Para la puesta en práctica del método de valoración de expertos se utilizaron las siguientes etapas,

seleccionadas entre las propuestas por (Hernández, 2000) teniéndose en cuenta las posibilidades reales para la validación de éste software.

- Selección de los posibles expertos
- Obtención del criterio de cada experto
- Procesamiento de los criterios de los expertos

Para el caso particular de la propuesta estas etapas se pusieron en práctica de la forma siguiente:

- Selección de los posibles expertos

  Según los criterios de (Cardona, 2003), se emplearon tres grupos de expertos evaluadores: el de contenido, el de diseño instruccional y el técnico informático. El experto en contenido tendrá la tarea de evaluar los aspectos de la ciencia correspondiente a la información que contempla el software; el experto en diseño instruccional será el encargado de evaluar los aspectos relacionados con la manera de presentar y enlazar la información; mientras que el experto en informática evaluará los aspectos relacionados con la interfaz, usabilidad, etc. A partir de la valoración subjetiva de los expertos se determinaron cuáles debían ser las variables adecuadas para someter a valoración el software.

  Para la selección de los posibles expertos se tuvieron en cuenta algunas variables propuestas en la bibliografía consultada, así como las características específicas del software que se valida.

  Se consideraron posibles expertos en contenido a aquellos profesores pertenecientes a la disciplina Matemática Aplicada que hayan impartido la asignatura Estadística Matemática más de 5 veces, como expertos en diseño instruccional a los profesores que hayan realizado al menos 3 trabajos relacionados con la temática y, además, ser un profesional relacionado con la psicopedagogía. Los expertos en informática serán los profesores ó no de la especialidad de Informática vinculados a la temática, ya sea en la programación o el diseño gráfico.

  La población de expertos fue seleccionada de los Departamentos de Matemática e Informática, pertenecientes a la Facultad de Informática y del departamento de Psicología de la Facultad de Ciencias Sociales y Humanísticas de la Universidad de Cienfuegos ¨Carlos R. Rodríguez¨. Además se solicitó la cooperación de profesores afines a la materia, de la Facultad de Ciencias Médicas de Cienfuegos Raúl Dorticós Torrado”.

A partir de los variables anteriores y de entrevistas realizadas se seleccionaron 9 posibles expertos en el contenido que ofrece el software, 3 en diseño instruccional y 15 posibles expertos en informática. Además, se tuvieron en cuenta otros criterios subjetivos tales como la creatividad, el interés por participar y la competencia del experto. Dichos resultados pueden observarse en los **Anexos 3A, 3B y 3C**.

- Obtención del criterio de cada experto

  Para la obtención de las valoraciones de los expertos primeramente se seleccionaron las dimensiones y variables para lo cual se tuvieron en cuenta los trabajos referenciados por (González, 2003), (Cardona, 2003) y del Grupo de Expertos del Ministerio de Educación de la República de Cuba (2004). Teniendo en cuenta estos trabajos y los resultados obtenidos según criterio de expertos, han sido seleccionadas las dimensiones y variables para la valoración del software, entre las cuales se encuentran:

  Para orientar la actividad valorativa del experto en contenido se incluyen dentro de esta dimensión, las siguientes variables:

  - Redacción
  - Lenguaje adecuado al nivel de enseñanza
  - Vigencia científica
  - Confiabilidad conceptual
  - Pertinencia
  - Transferencia de aprendizaje
  - Motivación
  - Confiabilidad psicopedagógica
  - Ayudas

  Para orientar la actividad valorativa del experto en diseño instruccional se incluyen dentro de esta dimensión, variables como:

  - Redacción
  - Presentación del software

- Textos
- Objetivos
- Secuencia lógica
- Flexibilidad
- Pertinencia
- Enfoque de aprendizaje
- Motivación
- Interacción
- Ayudas

Para orientar la actividad valorativa del experto en informática se incluyen dentro de esta dimensión, variables como:

- Estructura de programación
- Facilidad de comprensión
- Adaptabilidad
- Interfaz gráfica
- Confiabilidad funcional

Luego, con la finalidad de obtener los criterios valorativos de los expertos por dimensiones, a éstos se les entregaron los instrumentos que permitirían valorar el software **(Anexos 3D, 3E, 3F)**.

- Procesamiento del criterio de los expertos

Los resultados del procesamiento estadístico de las encuestas aplicadas acerca de la valoración subjetiva de los expertos se encuentran en los **Anexos 3G, 3H y 3I**.

Se encuestaron un total de 9 expertos que emitieron sus valoraciones sobre el contenido del software, los cuales otorgaron puntuaciones promedio de 4 puntos (total acuerdo) con un pequeño rango de variación de 0 a 1 puntos.

Los expertos en la **dimensión contenido** estuvieron en total acuerdo con los variables vigencia científica y motivación. Sobre el lenguaje adecuado al nivel de enseñanza y la

confiabilidad conceptual, el 88.9 % de los expertos consultados manifiestan total acuerdo mientras que el 11.1 % restante están de acuerdo. Por otra parte se manifiestan en total acuerdo con respecto a la redacción, pertinencia y transferencia de aprendizaje el 77.8 %, el 66.7 % y el 55.6 % respectivamente. El 33.3 % y un 22.2 % no emitieron criterios de la confiabilidad psicopedagógica y la ayuda respectivamente, aunque la mayoría restante se manifiesta en total acuerdo con el tratamiento dado a estas variables.

Es relevante señalar que en ningún caso se encontraron criterios de expertos en desacuerdo o total desacuerdo con los variables objeto de valoración.

Tabla 3.8: Fuente: Anexo 3G

Variables		C1	C2	C3	C4	NA	Total
X1	Redacción			2	7		9
X2	Lenguaje adecuado al nivel de enseñanza			1	8		9
X3	Vigencia científica			0	9		9
X4	Confiabilidad conceptual			1	8		9
X5	Pertinencia			3	6		9
X6	Transferencia de aprendizaje			4	5		9
X7	Motivación			0	9		9
X8	Confiabilidad psicopedagógica			2	4	3	9
X9	Ayudas			1	6	2	9

Leyenda:

- C1- Total desacuerdo
- C2- En desacuerdo
- C3- De acuerdo
- C4- Total acuerdo
- NA- No aplica (No tiene criterios sobre el item que se analiza)

Las tablas de frecuencias correspondientes a la **dimensión diseño instruccional** permitieron arribar a las siguientes conclusiones:

Se encuestaron un total de 3 expertos que emitieron sus valoraciones sobre el contenido del software, los cuales otorgaron puntuaciones promedio entre 3 y 4 puntos (de acuerdo - total acuerdo) con un pequeño rango de variación de 0 a 1 puntos. Los expertos en el diseño instruccional estuvieron en total acuerdo con los variables redacción, presentación del software, objetivos, pertinencia, interacción y ayuda. Sobre los textos y la secuencia lógica, el 66,7 % de los expertos consultados manifiestan total acuerdo mientras que el 33,3 % restante están de acuerdo. Por otra parte se manifiestan en total acuerdo o de acuerdo con respecto a la enfoque de aprendizaje y motivación, el 66.6 % de los encuestados, un 33.3 % no emite criterios al respecto. El 66.7 % se manifiesta de acuerdo con la flexibilidad.

Al igual que en el análisis anterior se destaca que en ningún caso se encontraron criterios de expertos en desacuerdo o total desacuerdo con los variables objeto de valoración.

Tabla 3.9: Fuente: Anexo 3H

Variables		C1	C2	C3	C4	NA	Total
X1	Redacción			0	3		3
X2	Presentación del software			0	3		3
X3	Textos			1	2		3
X4	Objetivos			0	3		3
X5	Secuencia lógica			1	2		3
X6	Flexibilidad			2	1		3
X7	Pertinencia			0	3		3
X8	Enfoque de aprendizaje			1	1	1	3
X9	Motivación			1	1	1	3
X10	Interacción			0	3		3
X11	Ayudas			0	3		3

Las tablas de frecuencias correspondientes a la **dimensión informática** permitieron arribar a las siguientes conclusiones:

Se encuestaron un total de 15 expertos en informática, los cuales no emitieron total acuerdo en ninguna de las variables analizadas, aunque otorgaron puntuaciones promedios entre 3 y 4 puntos (de acuerdo - total acuerdo) a dichas variables con un rango de variación de 1 a 2 puntos.

Acerca de la facilidad de compresión y la confiabilidad funcional el 73.3 % de los encuestados se manifiesta en total acuerdo con el tratamiento dado a dichas variables. A la estructura de programación se le otorga la mayor puntuación por el 60 % de los encuestados, un 26 % manifiesta no tener criterio sobre la dicha variable y el restante 13.3 % se manifiesta de acuerdo con la misma. Por otra parte se manifiestan en total acuerdo con respecto a la adaptabilidad y la interfaz gráfica, el 40 % y el 46.7 % respectivamente. En ambas variables predomina la escala de evaluación: de acuerdo (60 % y 53.3 %) respectivamente.

Tabla 3.10: Fuente: Anexo 3I

Variables		C1	C2	C3	C4	NA	Total
X1	Estructura de programación			2	9	4	15
X2	Facilidad de comprensión			4	11		15
X3	Adaptabilidad			9	6		15
X4	Interfaz gráfica			8	7		15
X5	Confiabilidad funcional			4	11		15

Para cumplimentar el análisis anterior se realizó la Prueba no Paramétrica W. de Kendall con el objetivo de demostrar estadísticamente la posible existencia de acuerdo entre los avaluadores. Dicha prueba contrasta la hipótesis nula que plantea que no hay acuerdo contra la hipótesis alternativa en que sí se considera que hay acuerdo entre los avaluadores. Tomando como referencia un nivel de significación del $5\,\%$, si este es menor que la significación asintótica, entonces rechazamos Ho, de lo contrario aceptamos. Por otra parte los rangos obtenidos en dicha prueba permiten ordenar los criterios analizados según la importancia atribuida por los expertos.

Utilizando un nivel de significación del $5\,\%$ al comparar con la significación asintótica de los estadísticos calculados se obtuvo ($0,044$, $0,048$ y $0,007$), entonces puede concluirse que se acepta la hipótesis alternativa en los análisis realizados para los tres grupos de expertos por lo tanto, existe concordancia de criterios entre los mismos y los planteamientos analizados. **Anexos 3J**, **3K** y **3L**

3.7. Conclusiones Parciales

La realización del estudio de factibilidad del producto informático proyectó una cantidad significativa de beneficios tangibles e intangibles como resultado de su futura implantación en centros de estudios, empresas, etc. El sistema propuesto contribuye de forma positiva en el proceso de elaboración del modelo de encuesta y por consiguiente proporciona un ahorro considerable de recursos, ello evidencia la factibilidad económica que representa su posterior implementación. Una vez concluido el estudio de factibilidad del sistema, se estima un tiempo de 22 meses para su construcción por un hombre y su costo asciende a $\$4826{,}00$.

El desarrollo de la validación del software mostró resultados favorables a partir de la evaluación del coeficiente Alpha de Cronbach para corroborar la fiabilidad del cuestionario con la obtención de un valor superior al $0{,}799$. Con la realización de la Prueba no Paramétrica W. de Kendall se demostró estadísticamente la existencia de concordancia en los criterios de los expertos.

CONCLUSIONES

Como resultado de las investigaciones teóricas y experimentales de este trabajo se pueden establecer las siguientes conclusiones:

1. La vinculación de las tecnologías informáticas a los procedimientos de investigación que involucren el cálculo del tamaño de muestras tiene influencia positiva en el proceso de confección y puesta en práctica del modelo de encuesta.

2. La determinación del número de elementos de la muestra mediante la aplicación de los diseños muestrales probabilísticos, permite inferir con mayor grado de precisión los estadísticos de la población objeto de estudio.

3. El desarrollo de un producto de software que comprenda los diseños fundamentales de la teoría del muestreo proporciona una herramienta de notable utilidad que contribuye al adecuado desempeño del experimentador en la gestión del tamaño muestral.

4. La incorporación de adecuadas prácticas de muestreo en el quehacer investigativo de la Universidad de Cienfuegos Carlos Rafael Rodríguez posibilita elevar el nivel científico del personal que labora en el centro.

RECOMENDACIONES

- Implantar el sistema informático en la Universidad de Cienfuegos Carlos Rafael Rodríguez a partir del próximo curso escolar en septiembre del 2008, como punto de partida para su posterior extensión a otros centros de estudio.
- Para una mejor refinación de los resultados en la práctica se propone validar la efectividad del sistema, mediante la realización de pruebas comparativas en relación al nivel de precisión y la representatividad de los estadísticos expuestos.
- Considerar la posibilidad de incorporar al sistema una de base de datos para gestionar a mayor escala los datos y resultados correspondientes al desarrollo de investigaciones de carácter muestral.
- Desarrollar una versión mejorada del sistema que incorpore otros diseños muestrales, con la finalidad de consolidar una herramienta de propósito integrador respecto al cálculo de la teoría del muestreo.

BIBLIOGRAFÍA

Arias, R. M. (1995). *El método de encuestas por muestreo: Conceptos básicos.* Colección Síntesis Psicológica.

Arrondo, V. M. (1996). *Tamaño Óptimo de muestra en investigación por encuestas.* (http://fondosdigitales.us.es/publicthesis/743/15709.pdf)

Bonilla, G. (1996). *Métodos prácticos de inferencia estadística.* Fondo de Cultura Económica.

Cardona, N. (2003). *Producción de software educativo.* Venezuela : Ed. Fundación Bolivariana de Informática y Telemática.

Cochran, G. W. (1976). *Técnicas de muestreo.* Compañía Editorial Continental S.A.

González, H. B. R. (2003). *Criterios para evaluar software educativo.* (http://www.iespana.es/byrong/index.htm)

Hernández, G. F. (2000). *El criterio de especialistas y su aplicación en las investigaciones pedagógicas.* [s.e.].

Hines, W. W. (1998). *Probabiliad y estadística.* Ed.Wiley.

Jacobson, G. y. R., I.; Booch. (2000). *El proceso unificado de desarrollo de software.* Addison-Wesley, EUA.

Mendenhall, S. (1987). *Elementos de muestreo.* Grupo Editorioal Iberoamérica.

Méndez, I. R. (1975). *Estadística y método científico.* Comunicaciones Técnicas.

Silva, L. C. (1993). *Muestreo para la investigación en ciencias de la salud.* Madrid: Díaz Santos.

Vinelo, A. C. (1990). *Técnicas de muestreo.* Combinado Poligráfico Osvaldo Sánchez.

Weimer, R. C. (1998). *Estadística.* Madrid: Thomson Ed.

ANEXOS

Anexo 1: Tabla de Cochran

Respecto al tamaño de muestra necesario, un matemático de renombre llamado Cochran ha establecido para diferentes valores de p el tamaño de muestra que es necesario tomar para garantizar una satisfactoria aproximación de la distribución binomial a la normal. Esto se refleja en la siguiente tabla:

Si $\widehat{p}$ es:	n debe ser como mínimo
0.5	30
0.4 o 0.6	50
0.3 o 0.7	80
0.2 o 0.8	200
0.1 o 0.9	600
0.05 o 0.95	1400

Anexo 2: Prototipos

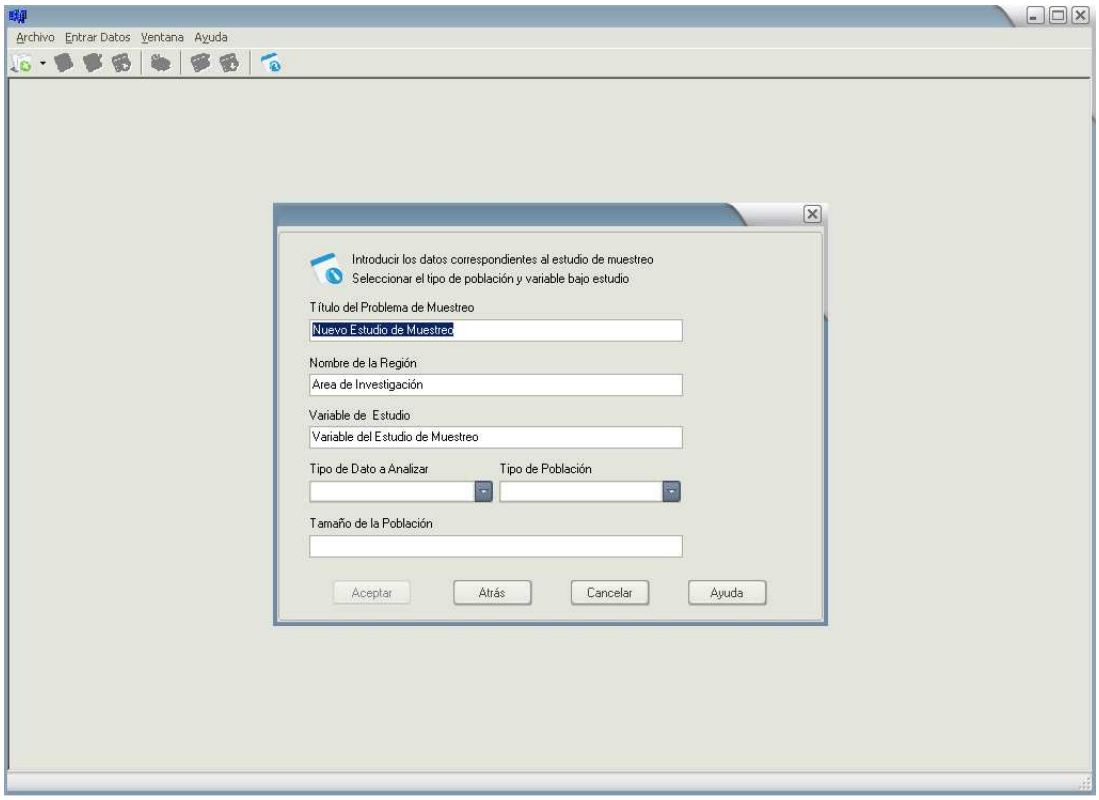


Figura 1: Gestionar los datos del estudio de muestreo

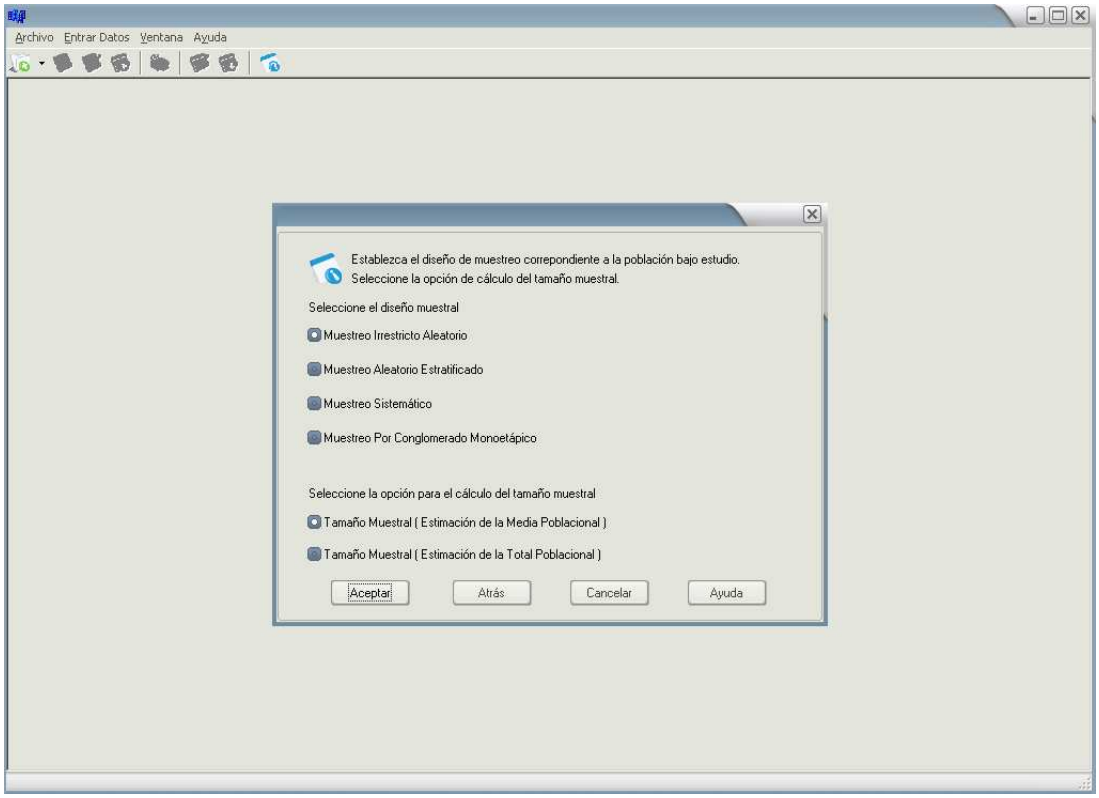


Figura 2: Gestionar los Datos Muestrales

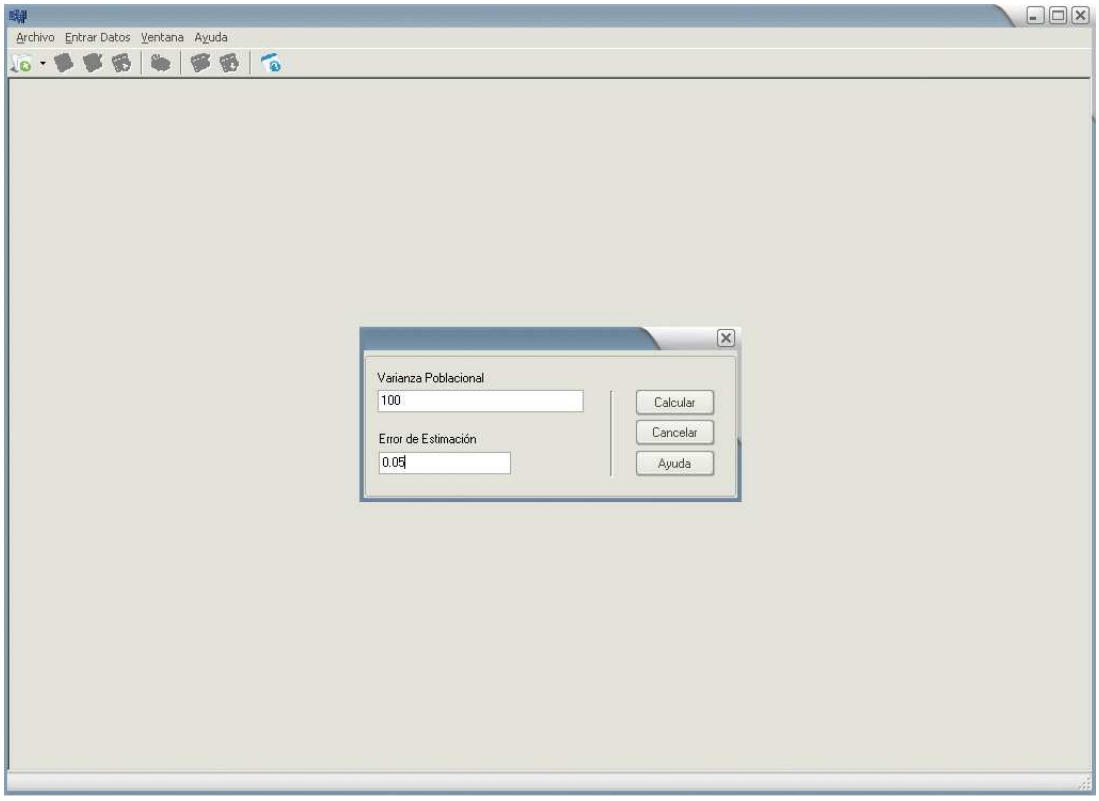
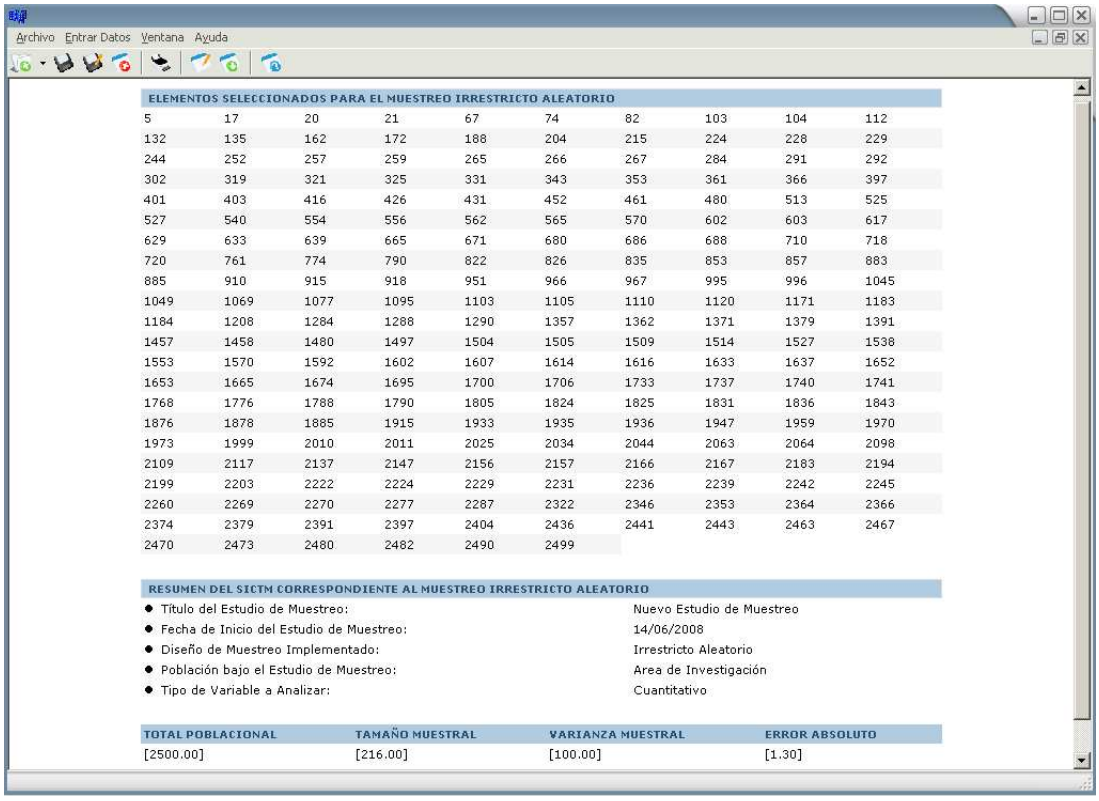


Figura 3: Confeccionar el reporte muestral (MIA)



Archivo Entrar Datos Ventana Ayuda

ELEMENTOS SELECCIONADOS PARA EL MUESTREO IRRESTRICTO ALEATORIO

5	17	20	21	67	74	82	103	104	112
132	135	162	172	188	204	215	224	228	229
244	252	257	259	265	266	267	284	291	292
302	319	321	325	331	343	353	361	366	397
401	403	416	426	431	452	461	480	513	525
527	540	554	556	562	565	570	602	603	617
629	633	639	665	671	680	686	688	710	718
720	761	774	790	822	826	835	853	857	883
885	910	915	918	951	966	967	995	996	1045
1049	1069	1077	1095	1103	1105	1110	1120	1171	1183
1184	1208	1284	1288	1290	1357	1362	1371	1379	1391
1457	1458	1480	1497	1504	1505	1509	1514	1527	1538
1553	1570	1592	1602	1607	1614	1616	1633	1637	1652
1653	1665	1674	1695	1700	1706	1733	1737	1740	1741
1768	1776	1788	1790	1805	1824	1825	1831	1836	1843
1876	1878	1885	1915	1933	1935	1936	1947	1959	1970
1973	1999	2010	2011	2025	2034	2044	2063	2064	2098
2109	2117	2137	2147	2156	2157	2166	2167	2183	2194
2199	2203	2222	2224	2229	2231	2236	2239	2242	2245
2260	2269	2270	2277	2287	2322	2346	2353	2364	2366
2374	2379	2391	2397	2404	2436	2441	2443	2463	2467
2470	2473	2480	2482	2490	2499				

RESUMEN DEL SICTM CORRESPONDIENTE AL MUESTREO IRRESTRICTO ALEATORIO

- Título del Estudio de Muestreo: Nuevo Estudio de Muestreo
- Fecha de Inicio del Estudio de Muestreo: 14/06/2008
- Diseño de Muestreo Implementado: Irrestricto Aleatorio
- Población bajo el Estudio de Muestreo: Area de Investigación
- Tipo de Variable a Analizar: Cuantitativo

TOTAL POBLACIONAL	TAMAÑO MUESTRAL	VARIANZA MUESTRAL	ERROR ABSOLUTO
[2500.00]	[216.00]	[100.00]	[1.30]

Figura 4: Confeccionar el reporte muestral (MIA)

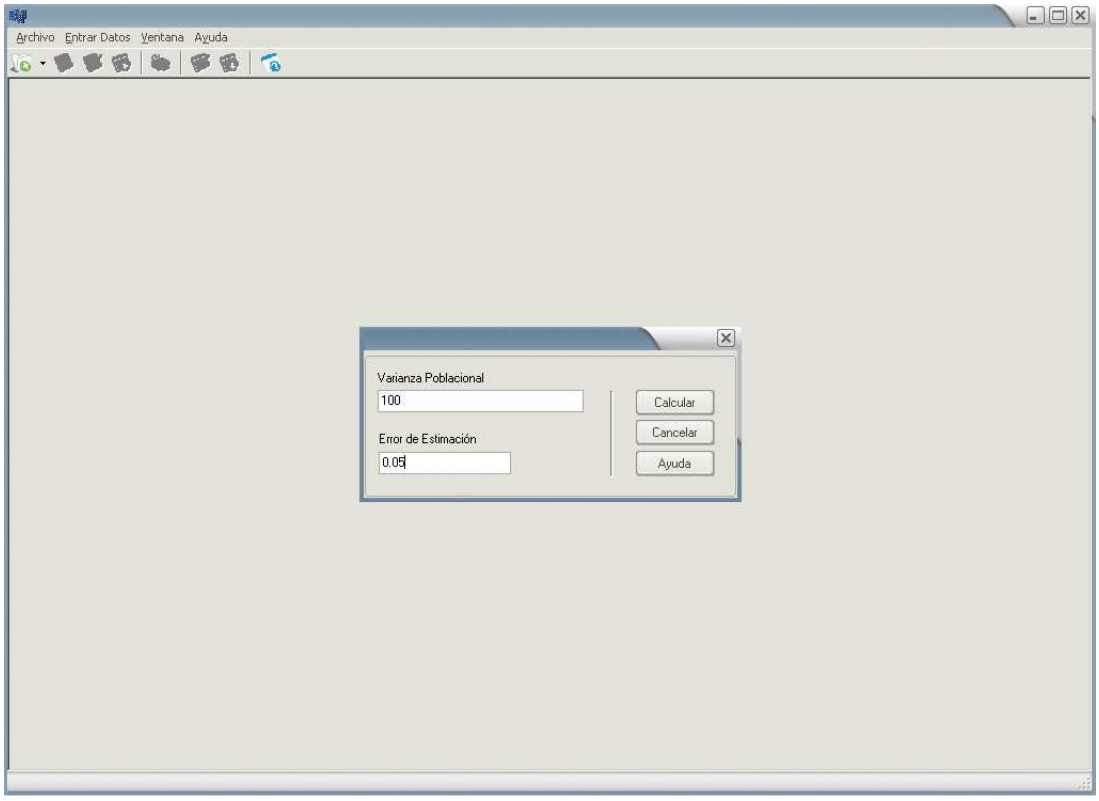
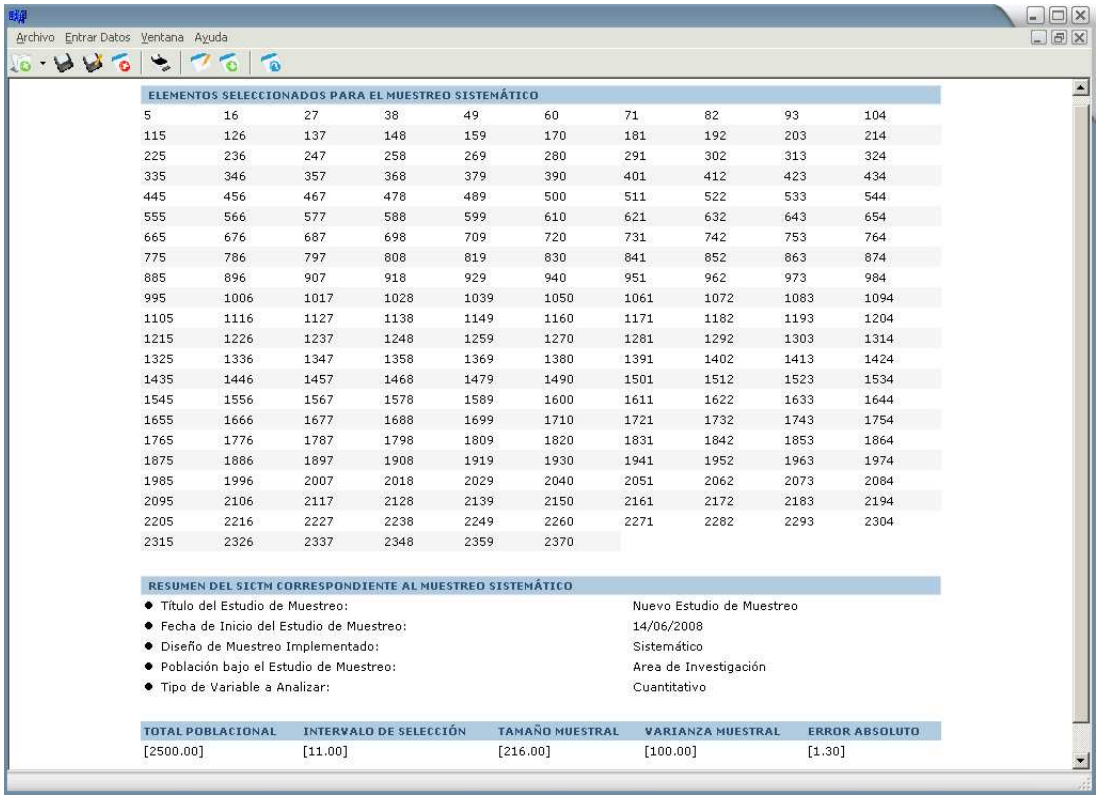


Figura 5: Confeccionar el reporte muestral (MSO)



Archivo Entrar Datos Ventana Ayuda

ELEMENTOS SELECCIONADOS PARA EL MUESTREO SISTEMÁTICO

5	16	27	38	49	60	71	82	93	104
115	126	137	148	159	170	181	192	203	214
225	236	247	258	269	280	291	302	313	324
335	346	357	368	379	390	401	412	423	434
445	456	467	478	489	500	511	522	533	544
555	566	577	588	599	610	621	632	643	654
665	676	687	698	709	720	731	742	753	764
775	786	797	808	819	830	841	852	863	874
885	896	907	918	929	940	951	962	973	984
995	1006	1017	1028	1039	1050	1061	1072	1083	1094
1105	1116	1127	1138	1149	1160	1171	1182	1193	1204
1215	1226	1237	1248	1259	1270	1281	1292	1303	1314
1325	1336	1347	1358	1369	1380	1391	1402	1413	1424
1435	1446	1457	1468	1479	1490	1501	1512	1523	1534
1545	1556	1567	1578	1589	1600	1611	1622	1633	1644
1655	1666	1677	1688	1699	1710	1721	1732	1743	1754
1765	1776	1787	1798	1809	1820	1831	1842	1853	1864
1875	1886	1897	1908	1919	1930	1941	1952	1963	1974
1985	1996	2007	2018	2029	2040	2051	2062	2073	2084
2095	2106	2117	2128	2139	2150	2161	2172	2183	2194
2205	2216	2227	2238	2249	2260	2271	2282	2293	2304
2315	2326	2337	2348	2359	2370				

RESUMEN DEL SICTM CORRESPONDIENTE AL MUESTREO SISTEMÁTICO

- Título del Estudio de Muestreo: Nuevo Estudio de Muestreo
- Fecha de Inicio del Estudio de Muestreo: 14/06/2008
- Diseño de Muestreo Implementado: Sistemático
- Población bajo el Estudio de Muestreo: Area de Investigación
- Tipo de Variable a Analizar: Cuantitativo

TOTAL POBLACIONAL	INTERVALO DE SELECCIÓN	TAMAÑO MUESTRAL	VARIANZA MUESTRAL	ERROR ABSOLUTO
[2500.00]	[11.00]	[216.00]	[100.00]	[1.30]

Figura 6: Confeccionar el reporte muestral (MSO)

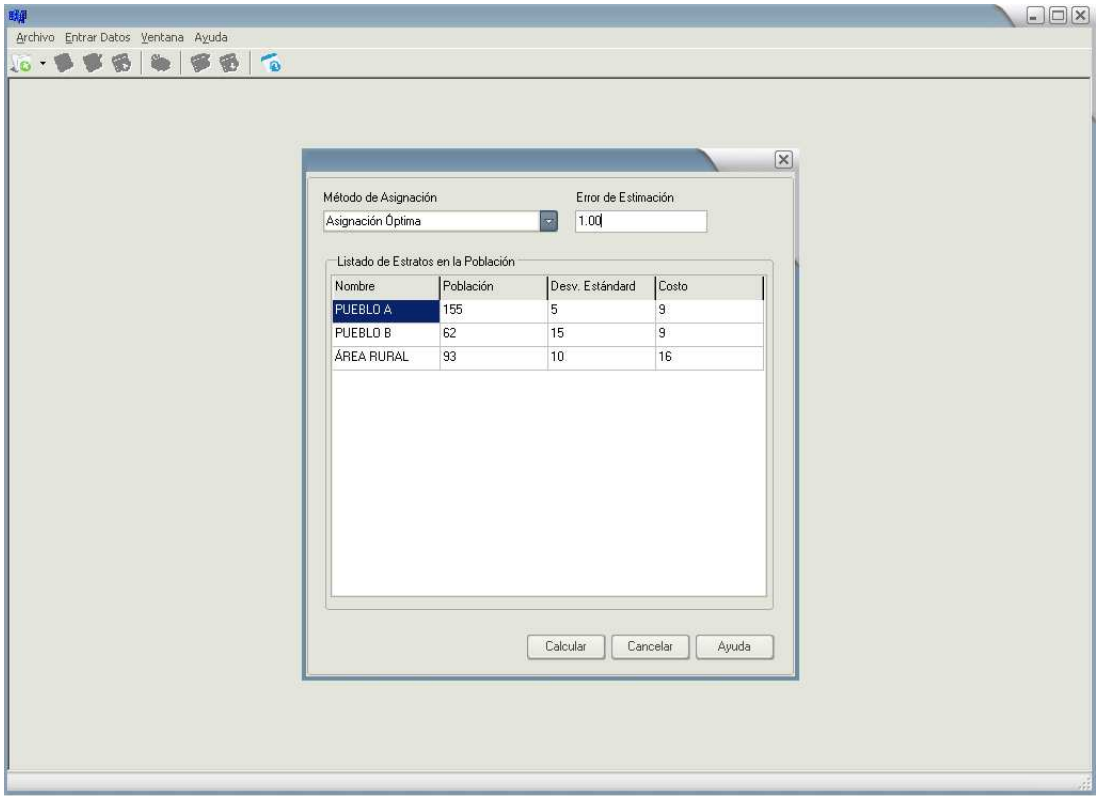
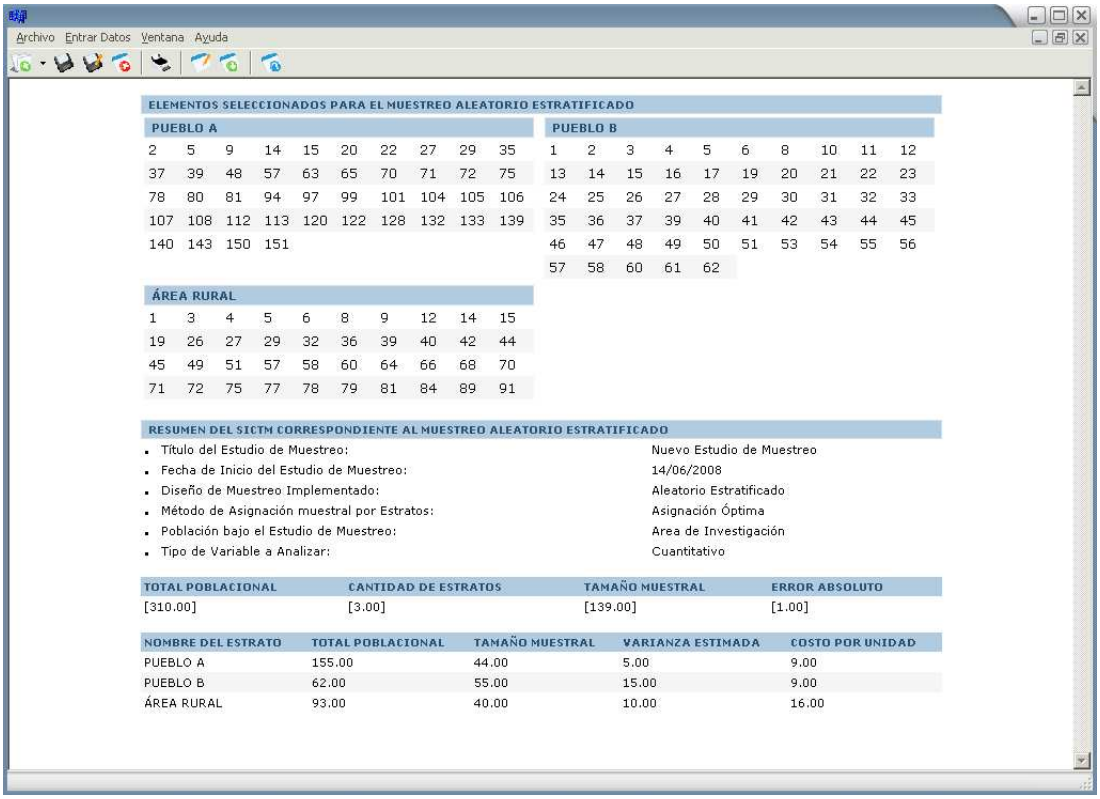


Figura 7: Gestionar los datos muestrales (MAE)



Archivo Entrar Datos Ventana Ayuda

ELEMENTOS SELECCIONADOS PARA EL MUESTREO ALEATORIO ESTRATIFICADO

PUEBLO A

2	5	9	14	15	20	22	27	29	35
37	39	48	57	63	65	70	71	72	75
78	80	81	94	97	99	101	104	105	106
107	108	112	113	120	122	128	132	133	139
140	143	150	151						

PUEBLO B

1	2	3	4	5	6	8	10	11	12
13	14	15	16	17	19	20	21	22	23
24	25	26	27	28	29	30	31	32	33
35	36	37	39	40	41	42	43	44	45
46	47	48	49	50	51	53	54	55	56
57	58	60	61	62					

ÁREA RURAL

1	3	4	5	6	8	9	12	14	15
19	26	27	29	32	36	39	40	42	44
45	49	51	57	58	60	64	66	68	70
71	72	75	77	78	79	81	84	89	91

RESUMEN DEL SICTM CORRESPONDIENTE AL MUESTREO ALEATORIO ESTRATIFICADO

- Título del Estudio de Muestreo: Nuevo Estudio de Muestreo
- Fecha de Inicio del Estudio de Muestreo: 14/06/2008
- Diseño de Muestreo Implementado: Aleatorio Estratificado
- Método de Asignación muestral por Estratos: Asignación Óptima
- Población bajo el Estudio de Muestreo: Area de Investigación
- Tipo de Variable a Analizar: Cuantitativo

TOTAL POBLACIONAL	CANTIDAD DE ESTRATOS	TAMAÑO MUESTRAL	ERROR ABSOLUTO
[310.00]	[3.00]	[139.00]	[1.00]

NOMBRE DEL ESTRATO	TOTAL POBLACIONAL	TAMAÑO MUESTRAL	VARIANZA ESTIMADA	COSTO POR UNIDAD
PUEBLO A	155.00	44.00	5.00	9.00
PUEBLO B	62.00	55.00	15.00	9.00
ÁREA RURAL	93.00	40.00	10.00	16.00

Figura 8: Confeccionar el reporte muestral(MAE)

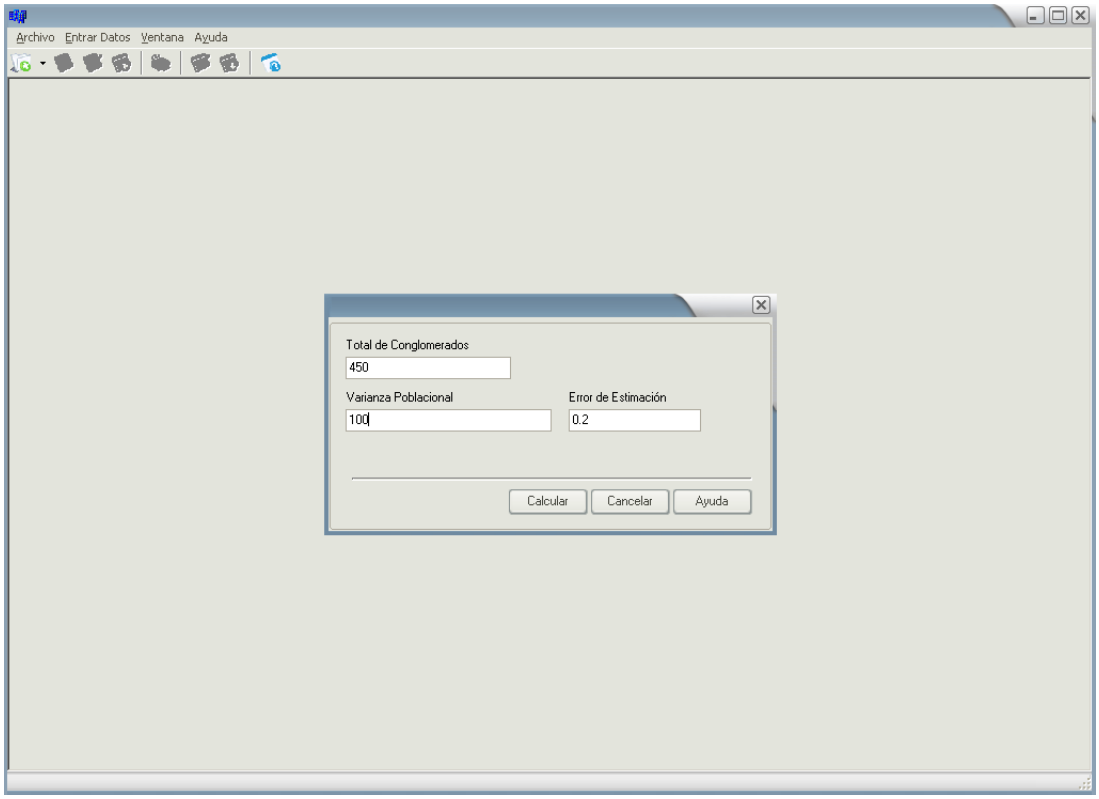
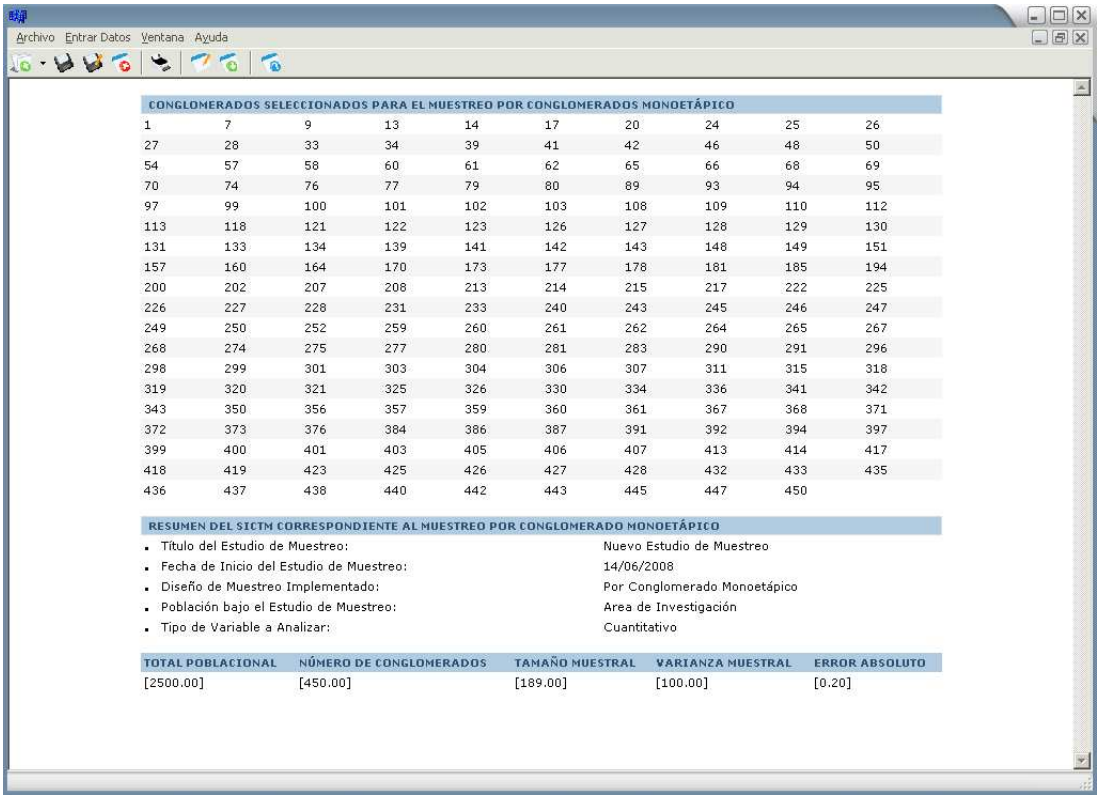


Figura 9: Confeccionar el reporte muestral(MCM)



Archivo Entrar Datos Ventana Ayuda

CONGLOMERADOS SELECCIONADOS PARA EL MUESTREO POR CONGLOMERADOS MONOETÁPICO

1	7	9	13	14	17	20	24	25	26
27	28	33	34	39	41	42	46	48	50
54	57	58	60	61	62	65	66	68	69
70	74	76	77	79	80	89	93	94	95
97	99	100	101	102	103	108	109	110	112
113	118	121	122	123	126	127	128	129	130
131	133	134	139	141	142	143	148	149	151
157	160	164	170	173	177	178	181	185	194
200	202	207	208	213	214	215	217	222	225
226	227	228	231	233	240	243	245	246	247
249	250	252	259	260	261	262	264	265	267
268	274	275	277	280	281	283	290	291	296
298	299	301	303	304	306	307	311	315	318
319	320	321	325	326	330	334	336	341	342
343	350	356	357	359	360	361	367	368	371
372	373	376	384	386	387	391	392	394	397
399	400	401	403	405	406	407	413	414	417
418	419	423	425	426	427	428	432	433	435
436	437	438	440	442	443	445	447	450	

RESUMEN DEL SICTM CORRESPONDIENTE AL MUESTREO POR CONGLOMERADO MONOETÁPICO

- Título del Estudio de Muestreo: Nuevo Estudio de Muestreo
- Fecha de Inicio del Estudio de Muestreo: 14/06/2008
- Diseño de Muestreo Implementado: Por Conglomerado Monoetápico
- Población bajo el Estudio de Muestreo: Area de Investigación
- Tipo de Variable a Analizar: Cuantitativo

TOTAL POBLACIONAL	NÚMERO DE CONGLOMERADOS	TAMAÑO MUESTRAL	VARIANZA MUESTRAL	ERROR ABSOLUTO
[2500.00]	[450.00]	[189.00]	[100.00]	[0.20]

Figura 10: Confeccionar el reporte muestral(MCM)

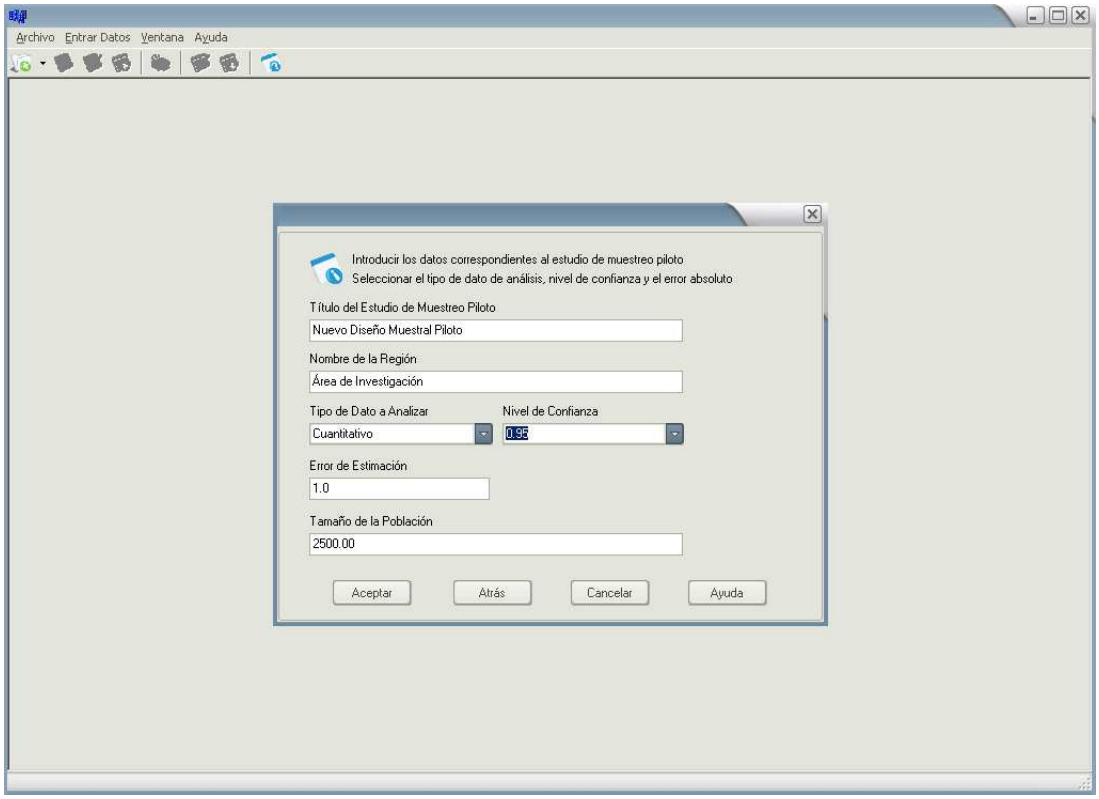


Figura 11: Gestionar los datos muestrales (MPL)

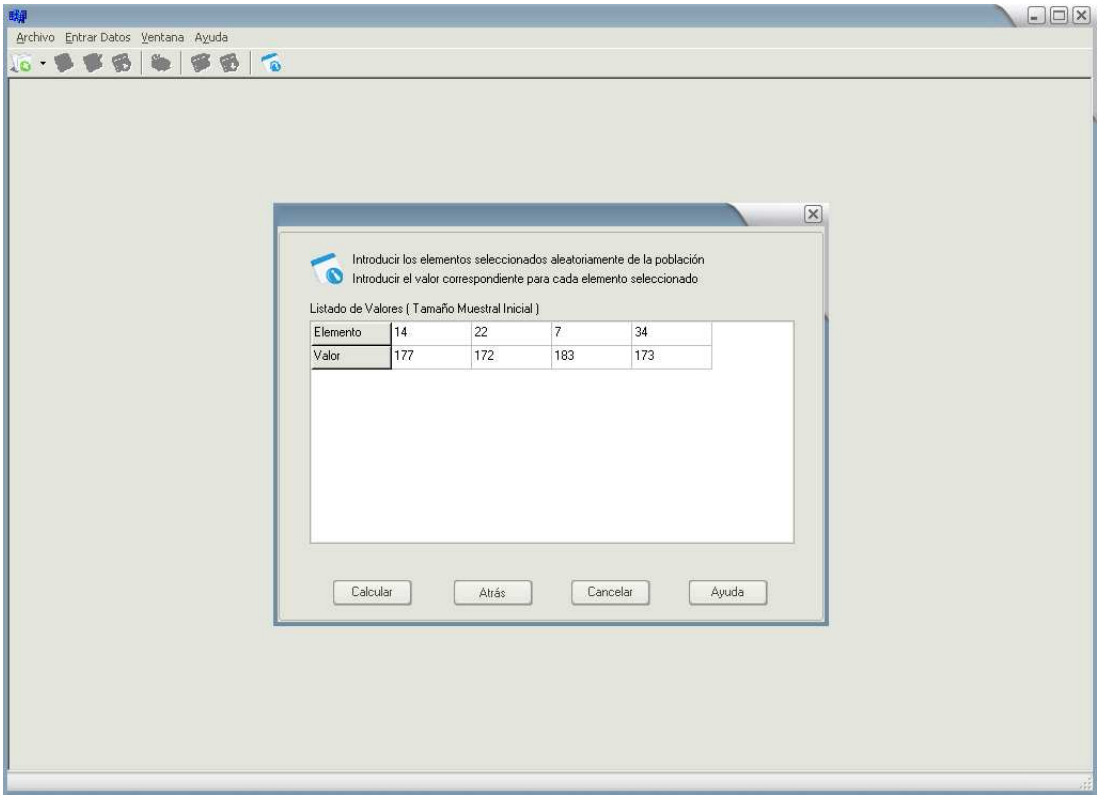
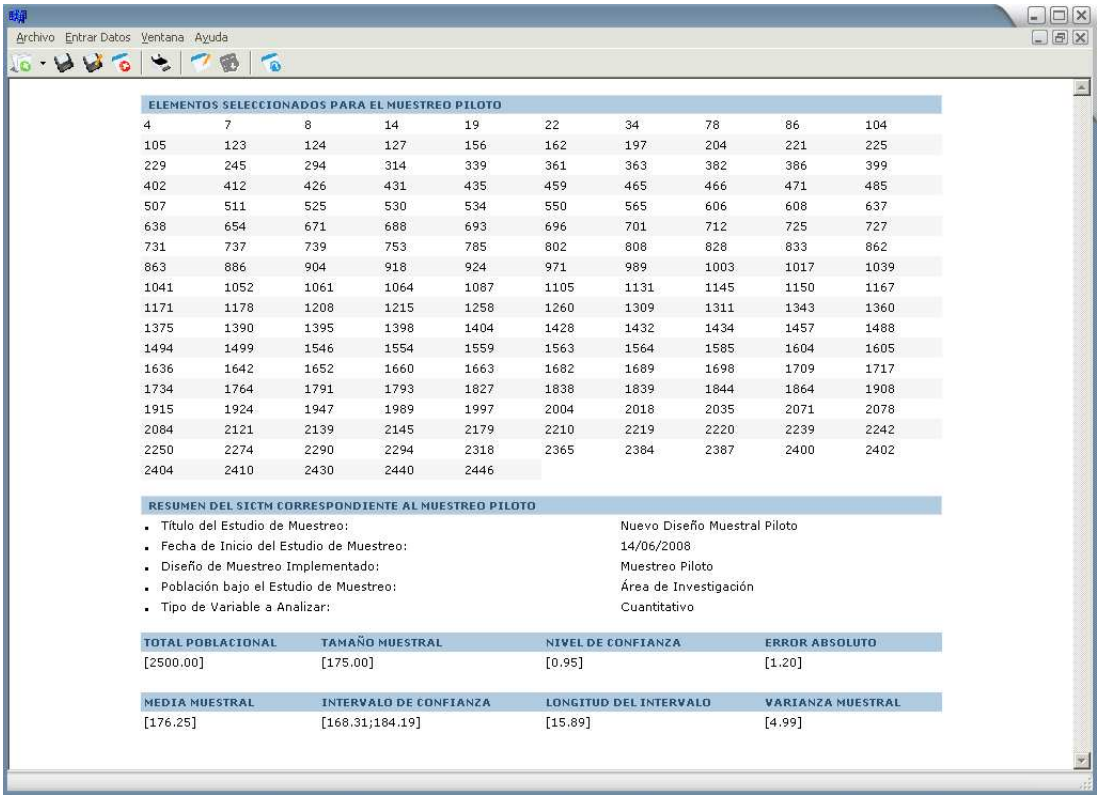


Figura 12: Confeccionar el reporte muestral(MPL)



Archivo Entrar Datos Ventana Ayuda

ELEMENTOS SELECCIONADOS PARA EL MUESTREO PILOTO

4	7	8	14	19	22	34	78	86	104
105	123	124	127	156	162	197	204	221	225
229	245	294	314	339	361	363	382	386	399
402	412	426	431	435	459	465	466	471	485
507	511	525	530	534	550	565	606	608	637
638	654	671	688	693	696	701	712	725	727
731	737	739	753	785	802	808	828	833	862
863	886	904	918	924	971	989	1003	1017	1039
1041	1052	1061	1064	1087	1105	1131	1145	1150	1167
1171	1178	1208	1215	1258	1260	1309	1311	1343	1360
1375	1390	1395	1398	1404	1428	1432	1434	1457	1488
1494	1499	1546	1554	1559	1563	1564	1585	1604	1605
1636	1642	1652	1660	1663	1682	1689	1698	1709	1717
1734	1764	1791	1793	1827	1838	1839	1844	1864	1908
1915	1924	1947	1989	1997	2004	2018	2035	2071	2078
2084	2121	2139	2145	2179	2210	2219	2220	2239	2242
2250	2274	2290	2294	2318	2365	2384	2387	2400	2402
2404	2410	2430	2440	2446					

RESUMEN DEL SICTM CORRESPONDIENTE AL MUESTREO PILOTO

- Título del Estudio de Muestreo: Nuevo Diseño Muestral Piloto
- Fecha de Inicio del Estudio de Muestreo: 14/06/2008
- Diseño de Muestreo Implementado: Muestreo Piloto
- Población bajo el Estudio de Muestreo: Área de Investigación
- Tipo de Variable a Analizar: Cuantitativo

TOTAL POBLACIONAL	TAMAÑO MUESTRAL	NIVEL DE CONFIANZA	ERROR ABSOLUTO
[2500.00]	[175.00]	[0.95]	[1.20]

MEDIA MUESTRAL	INTERVALO DE CONFIANZA	LONGITUD DEL INTERVALO	VARIANZA MUESTRAL
[176.25]	[168.31;184.19]	[15.89]	[4.99]

Figura 13: Confeccionar el reporte muestral(MPL)

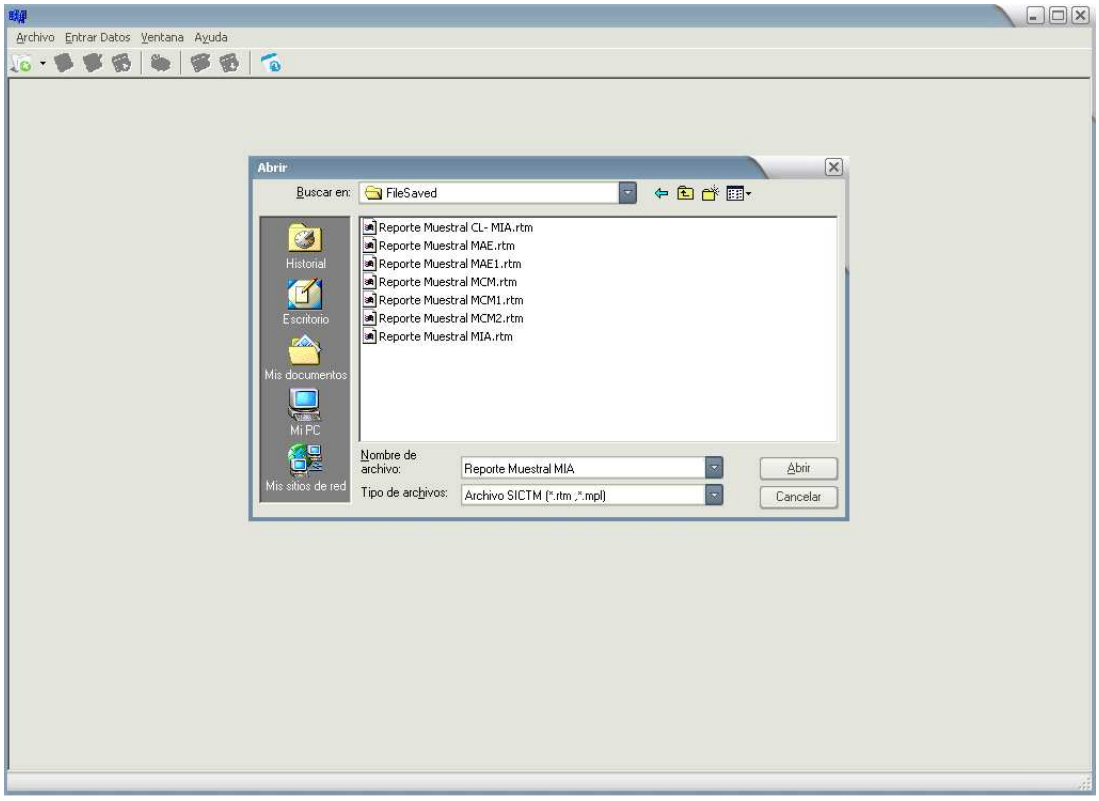
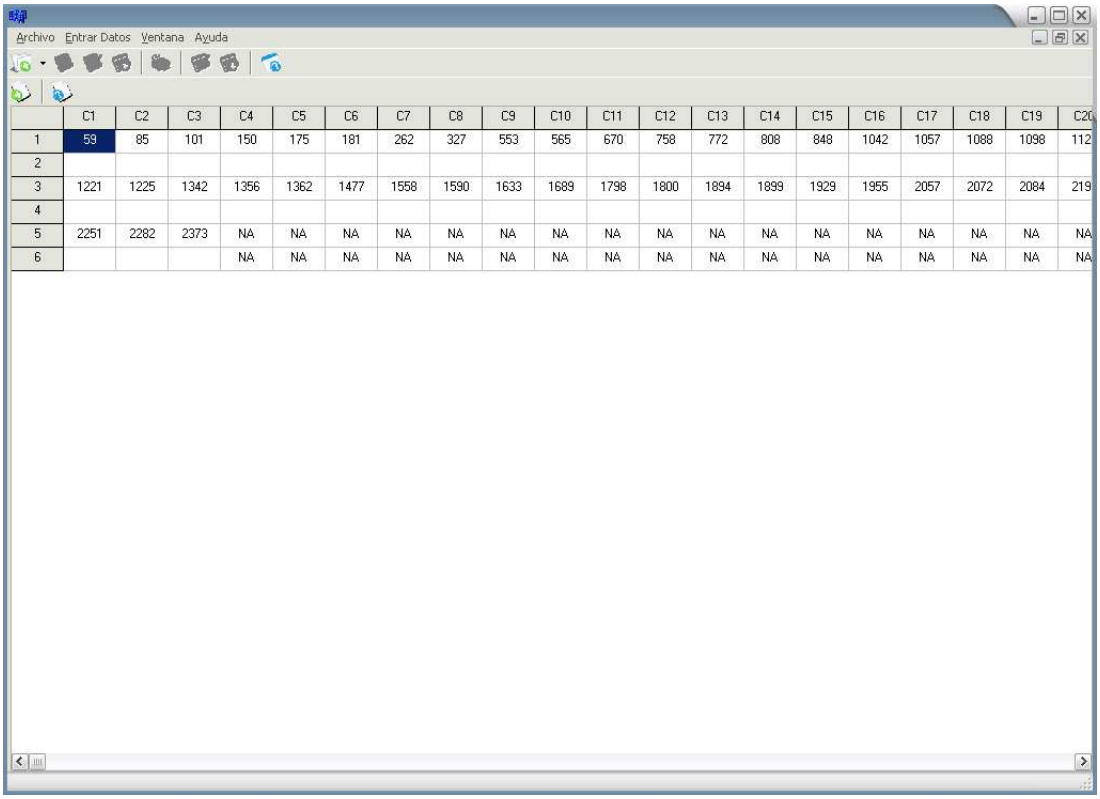


Figura 14: Insertar los datos muestrales



	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
1	59	85	101	150	175	181	262	327	553	565	670	758	772	808	848	1042	1057	1088	1098	112
2																				
3	1221	1225	1342	1356	1362	1477	1558	1590	1633	1689	1798	1800	1894	1899	1929	1955	2057	2072	2084	219
4																				
5	2251	2282	2373	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
6				NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Figura 15: Insertar los datos muestrales

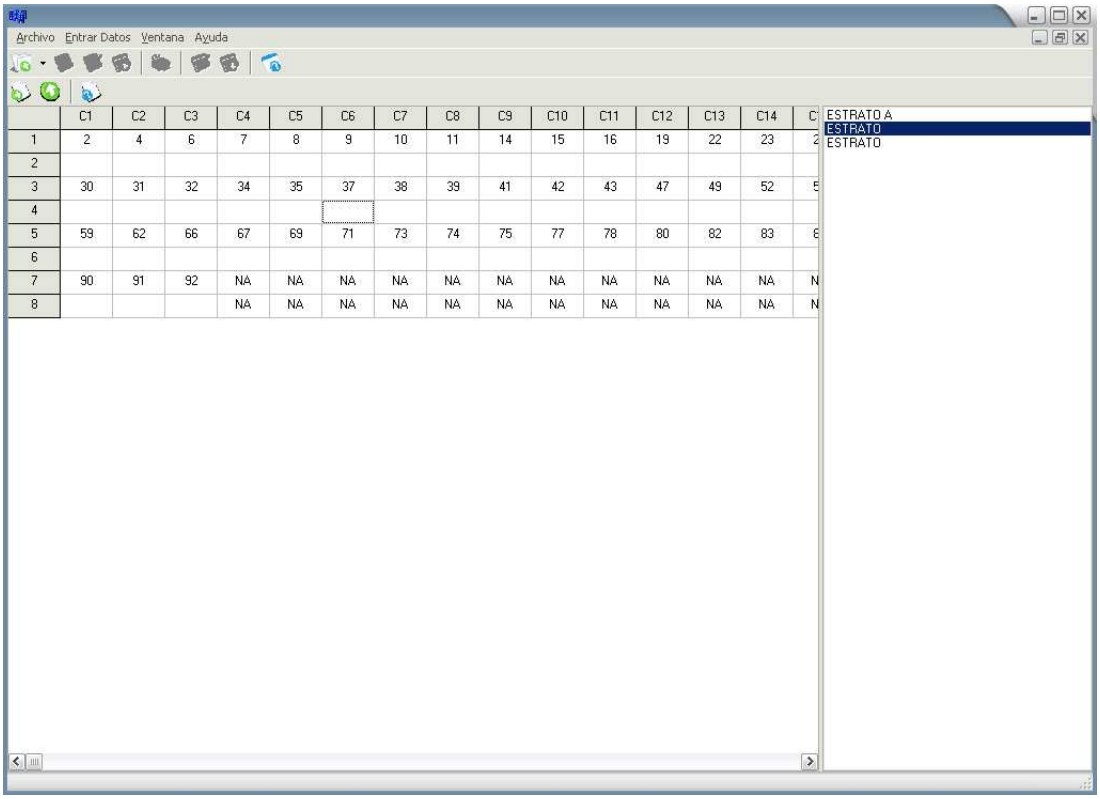


Figura 16: Insertar los datos muestrales

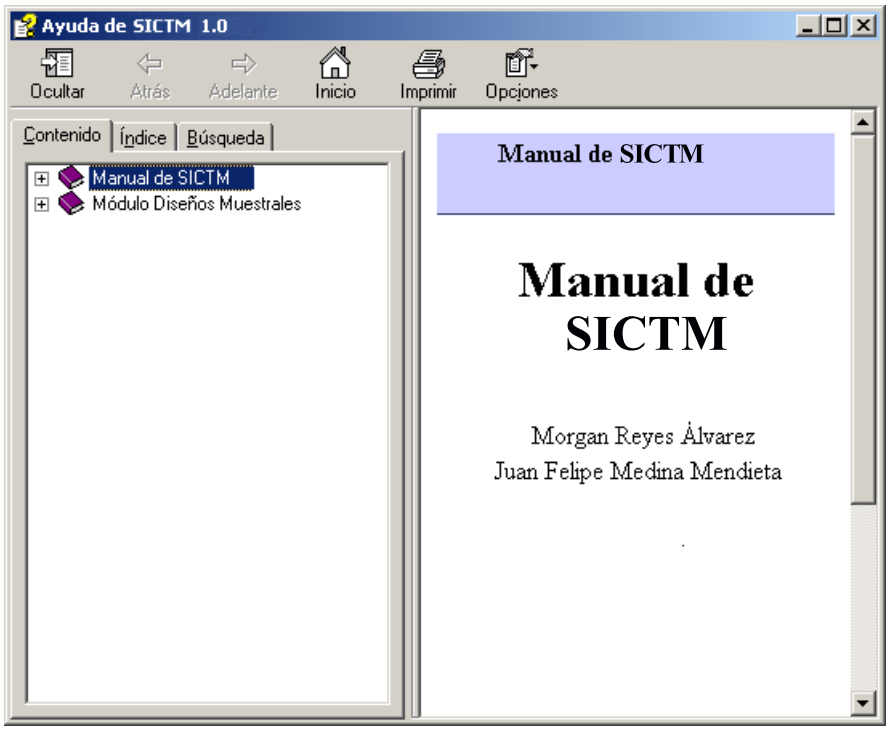


Figura 17: Consultar la ayuda del sistema

Anexo 3A. Características de los expertos en contenido

Número del experto	Años de experiencia	Categoría docente	Categoría científica	Nro de veces que ha impartido la asignatura	Trabajos investigativos realizados
1	30	Titular	Dr.	6	21
2	20	Auxiliar	MSc	36	11
3	25	Auxiliar	MSc	30	18
4	5	Instructor	-	7	2
5	15	Asistente	MSc	6	12
6	16	Auxiliar	MSc	12	9
7	10	Asistente	MSc	11	9
8	22	Auxiliar	MSc	10	10
9	30	Auxiliar	MSc	8	15

Anexo 3B. Características de los expertos en diseño instruccional

Número del experto	Años de experiencia	Categoría docente	Categoría científica	Relacionado con la psicopedagogía	Trabajos investigativos realizados
1	15	Auxiliar	MSc.	Sí	10
2	7	Asistente	MSc	Sí	5
3	4	Asistente	MSc	Sí	3

Anexo 3C. Características de los expertos en informática

Número del experto	Años de experiencia	Categoría docente	Categoría científica	Relacionado con la asignatura informática	Trabajos investigativos realizados
1	25	Asistente	MSc	Sí	15
2	26	Auxiliar	MSc	Sí	10
3	6	Asistente	-	Sí	2
4	22	Auxiliar	Msc	Sí	18
5	3	Instructor	MSc	Sí	3
6	6	Asistente	-	Sí	6
7	22	Auxiliar	MSc	Sí	18
8	3	Instructor	-	Sí	3
9	23	Asistente	MSc	Sí	7
10	9	Asistente	MSc	Sí	7
11	22	Auxiliar	MSc	Sí	15
12	5	Instructor	-	Sí	2
13	21	-	-	No	15
14	13	-	MSc	Sí	15
15	8	-	-	Sí	15

Anexo 3D. Encuesta aplicada a los expertos en el contenido para validar el Software SICTM

Estimado profesor, usted ha sido seleccionado como experto para ofrecer sus criterios valorativos acerca del contenido mostrado en el Software SICTM que permite calcular tamaños de muestras según los diferentes diseños existentes. Los investigadores le agradecen por su cooperación. Instrucciones: Para llenar el instrumento de evaluación es importante que siga los siguientes pasos:

- Evalúe las variables utilizando para ellos los variables que se adjuntan a cada una.
- Marque con una (X) en la escala de evaluación que se adjunta a cada variable utilizando la siguiente leyenda

  1. Total desacuerdo
  2. En desacuerdo
  3. De acuerdo
  4. Total acuerdo

  NA. No aplica. (Cuando el experto no tiene elementos suficientes para emitir un criterio de valor sobre el ítem).

- Cuando lo considere pertinente escriba sus criterios en la celda correspondiente a las Observaciones.

Lenguaje
REDACCIÓN 1 2 3 4 NA

a) La expresión de las ideas es clara y precisa.

b) Las estructuras gramaticales se utilizan correctamente.

Observaciones:
LENGUAJE ADECUADO AL NIVEL DE ENSEÑANZA 1 2 3 4 NA

a) El lenguaje utilizado requiere de un conocimiento elemental de Estadística.

b) El lenguaje es comprensible de acuerdo al nivel de enseñanza.

Observaciones:

Contenido
VIGENCIA CIENTÍFICA 1 2 3 4 NA

a) Los términos utilizados tienen validez científica.

b) Se observa coherencia de los objetivos del software con el contenido

c) Estimula el estudio independiente.

d) Proporciona rapidez y precisión en los cálculos realizados.

Observaciones:

CONFIABILIDAD CONCEPTUAL 1 2 3 4 NA

a) Los conceptos presentados en el software tienen rigor y precisión.

b) Los enfoques tratados en los diferentes diseños que se muestran provienen de fuentes bibliográficas prestigiosas y reconocidas

Observaciones:

PERTINENCIA 1 2 3 4 NA

a) La profundidad del contenido se adecua al usuario del producto.

b) El contenido tratado contribuye a la reflexión y al desarrollo de valores

Observaciones:

TRANSFERENCIA DE APRENDIZAJE 1 2 3 4 NA

a) El contenido es práctico y aplicable a todas las áreas del conocimiento.

b) Favorece el desarrollo del aprendizaje significativo.

c) Se vinculan conocimientos de otros temas estadísticos estudiados con anterioridad.

d) Favorece el desarrollo de la independencia cognoscitiva.

Observaciones:

Aspectos pedagógicos
MOTIVACIÓN 1 2 3 4 NA

a) Logra motivar por su manejabilidad.

b) La interactividad es apropiada para el usuario.

c) El contenido estimula su utilización.

d) Resulta confiable y elimina la consulta de tablas estadísticas para obtener el resultado.

Observaciones:

CONFIABILIDAD PSICOPEDAGÓGICA 1 2 3 4 NA

a) El programa es eficaz instructivamente.

b) El programa se adapta a las exigencias comunicativas del usuario.

c) Los estímulos visuales y auditivos permiten ampliar las reservas de representación del educando.

d) Propicia facilidades de percepción en cuanto a integridad, objetividad, constancia, selectividad y capacidad de observación.

e) Exige del estudiante un nivel de conocimientos teóricos y empírico

f) Favorece el interés.

Observaciones:

AYUDAS 1 2 3 4 NA

a Ofrece una guía sobre varios aspectos conceptuales.

b Vincula los elementos relacionados a la operabilidad del software.

Anexo 3E. Encuesta aplicada a los expertos en diseño instruccional para valorar el Software SICTM

Estimado profesor, usted ha sido seleccionado como experto para ofrecer sus criterios valorativos acerca del diseño instruccional del Software SICTM para calcular tamaños de muestras según los diferentes diseños existentes. Los investigadores agradecen por su cooperación.
Instrucciones: Para llenar el instrumento de evaluación es importante que siga los siguientes pasos:

- Evalúe las variables utilizando para ellos los variables que se adjuntan a cada una.
- Marque con una (X) en la escala de evaluación que se adjunta a cada variable utilizando la siguiente leyenda

  1. Total desacuerdo
  2. En desacuerdo
  3. De acuerdo
  4. Total acuerdo

  NA. No aplica. (Cuando el experto no tiene elementos suficientes para emitir un criterio de valor sobre el ítem).

- Cuando lo considere pertinente escriba sus criterios en la celda correspondiente a las Observaciones.

Comunicación
REDACCIÓN 1 2 3 4 NA

a) La expresión de las ideas planteadas se presenta en forma clara.

b) Las palabras que se utilizan son de significado preciso.

c) La construcción de las frases es correcta de acuerdo a las normas gramaticales.

d) La redacción de las ideas ofrece libertad de pensamiento al estudiante.

Observaciones:

PRESENTACIÓN DEL SOFTWARE 1 2 3 4 NA

a) Resulta atractiva de acuerdo con las características de la población para la cual fue diseñado el programa.

b) Se presentan instrucciones para el usuario.

c) Las instrucciones son claras.

Observaciones:

TEXTO 1 2 3 4 NA

a) Las pantallas presentan una cantidad apropiada de información.

b) El tamaño y tipo de letra permiten leer en forma rápida y comprensivamente.

c) La velocidad de presentación de los textos permite que, según las características del usuario, pueden ser leídos en su totalidad.

Observaciones:

OBJETIVOS 1 2 3 4 NA

a) Los objetivos son cubiertos con la información que se muestra a lo largo del software.

b) Están claramente definidos o se infieren fácilmente del material.

c) Se pueden adecuar al currículo.

d) Se adaptan a los estándares de pensamiento del estudiante según las teorías de desarrollo evolutivo.

e) Las estrategias y actividades son coherentes con los objetivos.

f) La forma de mostrarlo no es lineal (el usuario puede consultarlo por niveles).

g) Muestra los objetivos claramente o se infieren fácilmente.

Observaciones:

SECUENCIA LÓGICA 1 2 3 4 NA

a) El orden en que se presenta la información puede ser personalizada.

b) Las actividades están estructuradas con inicio desarrollo y cierre.

c) Los enlaces realizados entre las partes hacen que la estructura encaje perfectamente.

Observaciones:

FLEXIBILIDAD 1 2 3 4 NA

a) Permite la actualización de la información.

b) El programa se puede emplear en otros contextos diferentes para el que fue creado.

c) El profesor o el mismo alumno pueden adaptar el material a las necesidades de aprendizaje o entrenamiento.

Observaciones:

PERTINENCIA 1 2 3 4 NA

a) Los contenidos están adecuados al currículo.

b) Las herramientas audiovisuales propician la comprensión del contenido.

c) La cantidad de información por pantalla es adecuada.

Observaciones:

Diseño instruccional
ENFOQUE DE APRENDIZAJE 1 2 3 4 NA

a) Permite al alumno encontrar significado a través del material.

b) Da la posibilidad de construir y reconstruir el conocimiento.

c) Se activan los conocimientos previos de los estudiantes.

d) Logra establecer relaciones entre los conocimientos previos y los por aprender.

Observaciones:

MOTIVACIÓN 1 2 3 4 NA

a) Los mensajes son amigables, estimulantes y no agresivos.

b) El software resuelve de forma creativa, un conjunto de tareas para poder contribuir con efectividad a la solución de un problema dado.

c) El nivel de dificultad de la información es promedio para el nivel del usuario.

d) Las herramientas audiovisuales tienen capacidad para mantener la atención del estudiante.

e) El software revela la importancia práctica del conocimiento, plantea la tarea docente como un reto, garantiza una anticipación del logro efectivo de la tarea.

Observaciones:

INTERACCIÓN 1 2 3 4 NA

a) La estructuración del programa permite acceder sin dificultad a sus diferentes componentes (contenidos, actividades, niveles, servicios, etc).

b) El programa es sencillo, fácil de usar, amigable.

c) El software ofrece diferentes niveles y tipos de ayuda en dependencia de la complejidad situacional que se presente.

d) La interacción con el estudiante es la adecuada para el tipo de programa.

Observaciones:

AYUDAS 1 2 3 4 NA

a) Los mensajes de mensajes de ayuda son claros.

Observaciones:

DOCUMENTACIÓN 1 2 3 4 NA

a) Existe coherencia entre el programa y el material impreso que lo complementa.

Observaciones:

Anexo 3F. Encuesta aplicada a los expertos en informática para valorar la calidad técnica del Software SICTM

Estimado profesor, usted ha sido seleccionado como experto para ofrecer sus criterios valorativos acerca del diseño instruccional del Software SICTM para calcular tamaños de muestras según los diferentes diseños existentes. Los investigadores agradecen por su cooperación.
Instrucciones: Para llenar el instrumento de evaluación es importante que siga los siguientes pasos:

- Evalúe las variables utilizando para ellos los variables que se adjuntan a cada una.
- Marque con una (X) en la escala de evaluación que se adjunta a cada variable utilizando la siguiente leyenda

  1. Total desacuerdo
  2. En desacuerdo
  3. De acuerdo
  4. Total acuerdo

  NA. No aplica. (Cuando el experto no tiene elementos suficientes para emitir un criterio de valor sobre el ítem).
- Navegue por todo el software antes de comenzar a llenar el instrumento de valoración
- Cuando lo considere pertinente escriba sus criterios en la celda correspondiente a las Observaciones.

Programación
ESTRUCTURA DE PROGRAMACIÓN 1 2 3 4 NA

a) La estructura contempla el uso racional de variables sin cálculos innecesarios.

b) Posibilita la rehusabilidad del código.

c) La estructura es flexible, permitiendo la adición de otros módulos.

Observaciones:

Usabilidad
FACILIDAD DE COMPRENSIÓN 1 2 3 4 NA

a) Facilidad para consultar los comandos disponibles en el programa.

b) El diseño del software permite mantener orientado al usuario.

c) Cada módulo dispone de una opción de ayuda para el tópico en uso.

Observaciones:

ADAPTABILIDAD 1 2 3 4 NA

a) Es versátil en diversos contextos desde la perspectiva de su funcionalidad.

Observaciones:

INTERFAZ GRÁFICA 1 2 3 4 NA

a) El software muestra una interfaz innovadora.

b) El color empleado es adecuado.

c) Existe simplicidad, equilibrio, armonía, unidad.

d) Los botones, la distribución de información en las pantallas es consistente.

e) El tamaño y tipo de letras son adecuados.

f) La redundancia (verbal, textual, iconográfica, sensorial, etc.) que pueda existir, tiene sentido positivo.

g) La estructuración del programa permite acceder sin dificultades a sus principales componentes (contenidos, actividades, niveles, servicios, etc).

h) Los recursos empleados están armónicamente distribuidos (sin sobrecargar).

i) Hay facilidad de navegación entre las distintas pantallas.

Observaciones:

Funcionalidad

CONFIABILIDAD FUNCIONAL 1 2 3 4 NA

a) El software funciona correctamente en su ambiente.

b) Es adecuado el tiempo de respuesta a las acciones que realiza el usuario.

c) Emite resultados completos debido a que hay precisión en los códigos.

d) Estabilidad al presionar cualquier tecla que no esté asignada al software.

e) Las teclas cumplen con la función que se les ha asignado en el software.

Observaciones:

ANEXO 3G: Resultados estadísticos sobre las valoraciones de los expertos en contenido acerca del SICTM

Estadísticos

		Redacción	Lenguaje adecuado al nivel de enseñanza	Vigencia científica	Confiabilidad conceptual	Pertinencia	Transferencia de aprendizaje	Motivación	Confiabilidad psicopedagógica	Ayudas
N	Válidos	9	9	9	9	9	9	9	9	9
	Perdidos	0	0	0	0	0	0	0	0	0
Mediana		4,00	4,00	4,00	4,00	4,00	4,00	4,00	4,00	4,00
Moda		4	4	4	4	4	4	4	4	4
Rango		1	1	0	1	1	1	0	2	2
Mínimo		3	3	4	3	3	3	4	3	3
Máximo		4	4	4	4	4	4	4	5	5

Redacción

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	2	22,2	22,2	22,2
	Total acuerdo	7	77,8	77,8	100,0
	Total	9	100,0	100,0	

Lenguaje adecuado al nivel de enseñanza

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	1	11,1	11,1	11,1
	Total acuerdo	8	88,9	88,9	100,0
	Total	9	100,0	100,0	

Vigencia científica

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Total acuerdo	9	100,0	100,0	100,0

Confiabilidad conceptual

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	1	11,1	11,1	11,1
	Total acuerdo	8	88,9	88,9	100,0
	Total	9	100,0	100,0	

Pertinencia

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	3	33,3	33,3	33,3
	Total acuerdo	6	66,7	66,7	100,0
	Total	9	100,0	100,0	

Transferencia de aprendizaje

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	4	44,4	44,4	44,4
	Total acuerdo	5	55,6	55,6	100,0
	Total	9	100,0	100,0	

Motivación

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Total acuerdo	9	100,0	100,0	100,0

Confiabilidad psicopedagógica

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	2	22,2	22,2	22,2
	Total acuerdo	4	44,4	44,4	66,7
	NA	3	33,3	33,3	100,0
	Total	9	100,0	100,0	

Ayudas

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	1	11,1	11,1	11,1
	Total acuerdo	6	66,7	66,7	77,8
	NA	2	22,2	22,2	100,0
	Total	9	100,0	100,0	

ANEXO 3H: Resultados estadísticos sobre las valoraciones de los expertos en el diseño instruccional acerca del SICTM

Estadísticos

		Redacción	Presentación del software	Textos	Objetivos	Secuencia lógica	Flexibilidad	Pertinencia	Enfoque de aprendizaje	Motivación	Interacción	Ayudas
N	Válidos	3	3	3	3	3	3	3	3	3	3	3
	Perdidos	0	0	0	0	0	0	0	0	0	0	0
Mediana		4,00	4,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00	4,00	4,00
Moda		4	4	4	4	4	3	4	3[a]	3[a]	4	4
Rango		0	0	1	0	1	1	0	2	2	0	0
Mínimo		4	4	3	4	3	3	4	3	3	4	4

a. Existen varias modas. Se mostrará el menor de los valores.

Redacción

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Total acuerdo	3	100,0	100,0	100,0

Presentación del software

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Total acuerdo	3	100,0	100,0	100,0

Textos

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	1	33,3	33,3	33,3
	Total acuerdo	2	66,7	66,7	100,0
	Total	3	100,0	100,0	

Objetivos

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Total acuerdo	3	100,0	100,0	100,0

Secuencia lógica

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	1	33,3	33,3	33,3
	Total acuerdo	2	66,7	66,7	100,0
	Total	3	100,0	100,0	

Flexibilidad

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	2	66,7	66,7	66,7
	Total acuerdo	1	33,3	33,3	100,0
	Total	3	100,0	100,0	

Pertinencia

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Total acuerdo	3	100,0	100,0	100,0

Enfoque de aprendizaje

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	1	33,3	33,3	33,3
	Total acuerdo	1	33,3	33,3	66,7
	NA	1	33,3	33,3	100,0
	Total	3	100,0	100,0	

Motivación

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	1	33,3	33,3	33,3
	Total acuerdo	1	33,3	33,3	66,7
	NA	1	33,3	33,3	100,0
	Total	3	100,0	100,0	

Interacción

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Total acuerdo	3	100,0	100,0	100,0

Ayudas

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Total acuerdo	3	100,0	100,0	100,0

ANEXO 3I: Resultados estadísticos sobre las valoraciones de los expertos en informática acerca del SICTM

Estadísticos

		Estructura de programación	Facilidad de comprensión	Adaptabilidad	Interfaz gráfica	Confiabilidad funcional
N	Válidos	15	15	15	15	15
	Perdidos	0	0	0	0	0
Mediana		4,00	4,00	3,00	3,00	4,00
Moda		4	4	3	3	4
Rango		2	1	1	1	1
Mínimo		3	3	3	3	3
Máximo		5	4	4	4	4

Estructura de programación

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	2	13,3	13,3	13,3
	Total acuerdo	9	60,0	60,0	73,3
	NA	4	26,7	26,7	100,0
	Total	15	100,0	100,0	

Facilidad de comprensión

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	4	26,7	26,7	26,7
	Total acuerdo	11	73,3	73,3	100,0
	Total	15	100,0	100,0	

Adaptabilidad

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	9	60,0	60,0	60,0
	Total acuerdo	6	40,0	40,0	100,0
	Total	15	100,0	100,0	

Interfaz gráfica

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	8	53,3	53,3	53,3
	Total acuerdo	7	46,7	46,7	100,0
	Total	15	100,0	100,0	

Confiabilidad funcional

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	De acuerdo	4	26,7	26,7	26,7
	Total acuerdo	11	73,3	73,3	100,0
	Total	15	100,0	100,0	

ANEXO 3J: Prueba W. de Kendall para expertos en CONTENIDO

Rangos

	Rango promedio
Redacción	4,50
Lenguaje adecuado al nivel de enseñanza	5,00
Vigencia científica	5,50
Confiabilidad conceptual	5,06
Pertinencia	4,00
Transferencia de aprendizaje	3,61
Motivación	5,50
Confiabilidad psicopedagógica	5,89
Ayudas	5,94

Estadísticos de contraste

N	9
W de Kendall[a]	,219
Chi-cuadrado	13,685
gl	8
Sig. asintót.	,044

a. Coeficiente de concordancia de Kendall

ANEXO 3K: Prueba W. de Kendall para expertos en DISEÑO INSTRUCCIONAL

Rangos

	Rango promedio
Redacción	6,67
Presentación del software	6,67
Textos	4,83
Objetivos	6,67
Secuencia lógica	4,83
Flexibilidad	3,33
Pertinencia	6,67
Enfoque de aprendizaje	6,50
Motivación	6,50
Interacción	6,67
Ayudas	6,67

Estadísticos de contraste

N	3
W de Kendall[a]	,260
Chi-cuadrado	10,800
gl	10
Sig. asintót.	,048

a. Coeficiente de concordancia de Kendall

ANEXO 3L: Prueba W. de Kendall para expertos en INFORMÁTICA

Rangos

	Rango promedio
Estructura de programación	3,87
Facilidad de comprensión	3,13
Adaptabilidad	2,33
Interfaz gráfica	2,43
Confiabilidad funcional	3,23

Estadísticos de contraste

N	15
W de Kendall[a]	,233
Chi-cuadrado	13,951
gl	4
Sig. asintót.	,007

a. Coeficiente de concordancia de Kendall