

UNIVERSIDAD DE EL SALVADOR  
FACULTAD DE INGENIERIA Y ARQUITECTURA  
ESCUELA DE INGENIERIA DE SISTEMAS INFORMATICOS



**“Desarrollo e Integración de funciones estadísticas para el análisis de datos en el Sistema Integrado de Indicadores Gerenciales del Ministerio de Salud”**

PRESENTADO POR:

**WILLIAN ROBERTO GARCIA HERNANDEZ  
MAURICIO GIOVANNI GUERRERO CONTRERAS**

PARA OPTAR AL TITULO DE:

**INGENIERO DE SISTEMAS INFORMATICOS**

CIUDAD UNIVERSITARIA, MARZO DE 2016

**UNIVERSIDAD DE EL SALVADOR**

RECTOR INTERINO :

**LIC. JOSÉ LUIS ARGUETA ANTILLÓN**

SECRETARIO GENERAL :

**DRA. ANA LETICIA ZA VALETA DE AMAYA**

**FACULTAD DE INGENIERIA Y ARQUITECTURA**

DECANO :

**ING. FRANCISCO ANTONIO ALARCÓN SANDOVAL**

SECRETARIO :

**ING. JULIO ALBERTO PORTILLO**

**ESCUELA DE INGENIERIA DE SISTEMAS INFORMATICOS**

DIRECTOR :

**ING. JOSE MARIA SANCHEZ CORNEJO**

UNIVERSIDAD DE EL SALVADOR  
FACULTAD DE INGENIERIA Y ARQUITECTURA  
ESCUELA DE INGENIERIA DE SISTEMAS INFORMATICOS

Trabajo de Graduación previo a la opción al Grado de:

**INGENIERO DE SISTEMAS INFORMATICOS**

Título :

**“Desarrollo e Integración de funciones estadísticas para el análisis de datos en el Sistema Integrado de Indicadores Gerenciales del Ministerio de Salud”**

Presentado por :

**WILLIAN ROBERTO GARCIA HERNANDEZ  
MAURICIO GIOVANNI GUERRERO CONTRERAS**

Trabajo de Graduación Aprobado por :

Docente Asesor :

**ING. RODRIGO ERNESTO VASQUEZ ESCALANTE**

San Salvador, Marzo de 2016

Trabajo de Graduación Aprobado por:

Docente Asesor :

**ING. RODRIGO ERNESTO VASQUEZ ESCALANTE**

## **Agradecimientos**

Principalmente a **DIOS** por darme la oportunidad de culminar de forma exitosa este trabajo y brindarme paciencia, perseverancia, entrega y sobre todo por la salud que me ha regalado.

**A mi madre**, por estar pendiente de mí en todo momento, brindándome apoyo incondicional y por ser parte importante en el desarrollo del presente trabajo, su amor de madre, comprensión y ayuda ha sido fundamental para la finalización con éxito de mi carrera profesional.

**A mis amigos**, por la ayuda que siempre me han brindado. Por las alegrías y tristezas que hemos superado juntos, especialmente a **Sandra Carolina Pérez**.

**A mi compañero de tesis**, por nunca rendirse, siempre esforzarse hombro a hombro conmigo hasta lograr alcanzar esta meta.

**A nuestro docente asesor**, ingeniero Rodrigo Vásquez, por haber sido un guía formador, quien con su ayuda y apoyo nos permitiera que este trabajo de graduación sea un éxito.

*Willian Roberto García Hernández*

## **Agradecimientos**

En primer lugar, quiero agradecer a **DIOS** por permitirme culminar una nueva etapa de mi formación profesional, brindarme la sabiduría, paciencia y perseverancia necesaria para día con día avanzar en el proceso de la vida.

**A mis padres**, por estar en todo momento apoyándome y alentándome a alcanzar mis sueños y proyectos, por ayudarme y brindarme una educación integral que estoy seguro es el mejor legado que han podido obsequiarme.

**A mis hermanos y hermana**, por el apoyo y aliento que siempre me han brindado. Por las alegrías y tristezas que hemos superado juntos.

**A mi amiga, Sylvia**, por todo el apoyo, comprensión y motivación recibido que sirvió para nunca rendirme y lograr llevar a su culminación este trabajo de graduación.

**A mi compañero de tesis**, por nunca rendirse y siempre esforzarse hombro a hombro conmigo hasta lograr alcanzar esta meta.

**A nuestro docente asesor**, ingeniero Rodrigo Vásquez, por haber sido un guía formador, quien con su ayuda y apoyo nos permitiera que este trabajo de graduación sea un éxito.

*Mauricio Giovanni Guerrero Contreras*

# INDICE

INTRODUCCIÓN.....	10
OBJETIVOS .....	13
GENERAL.....	13
ESPECÍFICOS.....	13
ALCANCES .....	14
IMPORTANCIA.....	15
JUSTIFICACIÓN .....	16

## CAPÍTULO 1: INVESTIGACIÓN PRELIMINAR

1.1	EL ALMACÉN DE DATOS DEL MINISTERIO DE SALUD.....	19
1.1.1	ELEMENTOS QUE CONSTITUYEN LA ARQUITECTURA DE UN ALMACÉN DE DATOS .....	19
1.1.2	ESTRUCTURA DE UN ALMACÉN DE DATOS.....	21
1.1.3	ARQUITECTURA DEL ALMACÉN DE DATOS DEL MINISTERIO DE SALUD ...	23
1.1.4	DESCRIPCIÓN DE LA ARQUITECTURA DEL ALMACÉN DE DATOS DEL MINISTERIO DE SALUD .....	24
1.1.5	ESTRUCTURA DEL ALMACÉN DE DATOS DEL MINISTERIO DE SALUD .....	25
1.2	LOS PROCESOS EN ALMACENES DE DATOS .....	26
1.2.1	REPRESENTACIÓN DE PROCESO ETL EN ALMACÉN DE DATOS DEL MINISTERIO DE SALUD .....	27

1.3	EL ANÁLISIS DE DATOS .....	30
1.3.1	ANÁLISIS DE DATOS .....	30
1.3.2	EL CONCEPTO DE ANÁLISIS ESTADÍSTICO DE DATOS .....	31
1.4	LA INFORMÁTICA EN EL ANÁLISIS ESTADÍSTICO DE DATOS .....	33
1.5	PRINCIPALES FORMULAS ESTADÍSTICAS PARA EL ANÁLISIS DE DATOS.....	35
1.6	LOS INDICADORES ESTANDARIZADOS DE SALUD .....	37

## CAPÍTULO 2: SITUACIÓN ACTUAL

2.1	SITUACIÓN ACTUAL .....	40
2.1.1	DESCRIPCIÓN .....	40
2.1.2	REPRESENTACIÓN MEDIANTE ENFOQUE DE SISTEMAS .....	42
2.1.3	DESCRIPCIÓN DE LOS ELEMENTOS DEL ENFOQUE DE SISTEMAS.....	43

## CAPÍTULO 3: ANÁLISIS DEL PROBLEMA

3.1	PLANTEAMIENTO DEL PROBLEMA.....	48
3.1.1	MÉTODO DE LA CAJA NEGRA .....	50
3.1.2	DETALLE DEL PLANTEAMIENTO DEL PROBLEMA .....	51
3.2	FORMULACIÓN DEL PROBLEMA .....	52
3.3	ENFOQUE DE SISTEMAS PROPUESTO .....	53
3.4	DESCRIPCIÓN DE ELEMENTOS ENFOQUE DE SISTEMAS PROPUESTO .....	54

## CAPÍTULO 4: ANÁLISIS DE REQUERIMIENTOS

4.1	ANÁLISIS Y DETERMINACIÓN DE REQUERIMIENTOS .....	58
4.1.1	REQUERIMIENTOS FUNCIONALES .....	58
4.1.2	DESCRIPCIÓN DE REQUERIMIENTOS FUNCIONALES SOBRE FUNCIONES ESTADÍSTICAS .....	60
4.1.3	REQUERIMIENTOS NO FUNCIONALES .....	63
4.1.4	RESTRICCIONES PARA EL DESARROLLO .....	66
4.2	CASOS DE USO .....	69
4.2.1	CASO DE USO GENERAL .....	69
4.2.2	DESCRIPCIÓN CASOS DE USO .....	70
4.3	ELABORACIÓN DIAGRAMA ENTIDAD RELACIÓN .....	81
4.3.1	DIAGRAMA ENTIDAD RELACION DEL SISTEMA DE INDICADORES .....	82
4.3.2	DIAGRAMA ENTIDAD RELACIÓN MÓDULO DE ANALISIS DE DATOS .....	83
4.4	PROCEDIMIENTO PARA INTEGRAR FUNCIONES ESTADÍSTICAS .....	84

## CAPÍTULO 5: DISEÑO

5.1	DISEÑO DE LAS SALIDAS .....	101
5.1.1	MENÚ PRINCIPAL .....	101
5.1.2	MENÚ FUNCIONES ESTADÍSTICAS .....	101
5.1.3	MENÚ TABLERO .....	102
5.2	DISEÑO DE LAS FUNCIONES ESTADÍSTICAS .....	106
5.3	DISEÑO DE LOS PROCESOS DEL MÓDULO DE ANÁLISIS DE DATOS .....	128

5.4	DISEÑO DE LAS INTERFASES GRÁFICAS DE USUARIO.....	135
5.4.1	INTERFAZ GENERAL DEL SISTEMA INTEGRADO DE INDICADORES GERENCIALES .....	135

## CAPÍTULO 6: DESARROLLO DEL MÓDULO DE ANÁLISIS DE DATOS E IMPLEMENTACIÓN

6.1	ESTÁNDARES DE DESARROLLO.....	137
6.1.1	ESTÁNDARES DE LA BASE DE DATOS.....	137
6.1.2	ESTÁNDARES DE PROGRAMACIÓN.....	138
6.1.3	PATRÓN DE PROGRAMACIÓN MVC .....	139
6.2	FUNCIONES ESTADÍSTICAS DEL MÓDULO DE ANÁLISIS DE DATOS.....	141
6.2.1	FUNCIÓN MATRIZ DE CORRELACIONES .....	142
6.2.2	FUNCIÓN DENDOGRAMA .....	144
6.2.3	FUNCIÓN RESUMEN ESTADISTICO.....	147
6.2.4	FUNCIÓN CÍRCULO DE CORRELACION .....	149
6.2.5	FUNCIÓN CENTROS DE GRAVEDAD.....	152
6.2.6	FUNCIÓN INTERVALOS DE CONFIANZA .....	155
6.2.7	FUNCIÓN MAPA DE CORRELACION .....	157
6.2.8	FUNCIÓN ANALISIS DE CONGLOMERADOS .....	160
6.2.9	FUNCIÓN DIAGRAMA DE CAJA Y BIGOTE .....	163
6.2.10	FUNCIÓN DATOS ATÍPICOS O MAL REPRESENTADOS .....	166
6.2.11	FUNCIÓN REGRESIÓN LINEAL.....	168

6.3	INTEGRACIÓN CON EL SISTEMA DE INDICADORES .....	171
6.4	DOCUMENTACIÓN .....	172
6.5	PLAN DE IMPLEMENTACIÓN.....	173
	CONCLUSIONES.....	175
	RECOMENDACIONES .....	176
	GLOSARIO.....	177
	REFERENCIAS .....	181
	ANEXOS.....	184
	ANEXO 1 -INVESTIGACIÓN DETALLADA SOBRE FUNCIONES ESTADÍSTICAS .....	185
	ANEXO 2 -RESUMEN DE INDICADORES PRIORIZADOS Y ESTANDARIZADOS EN SALUD .....	202
	ANEXO 3 -SISTEMAS DE DESARROLLO Y PRODUCCION DEL MINISTERIO DE SALUD .....	205

## INTRODUCCIÓN

El Ministerio de Salud y Asistencia Social de El Salvador (MINSAL) es una institución gubernamental cuyos orígenes se remontan al año 1900. En sus inicios sus funciones principales eran: estadísticas médicas, saneamiento de zonas urbanas, inspecciones de víveres, higiene de rastros y mercados, construcción de cloacas y sistemas de aguas servidas, regir la obligatoriedad de instalar letrinas, lucha contra los mosquitos, visitas a establos, fábricas y beneficios de lavar café.

Con el paso del tiempo dichas funciones se han ido extendiendo y adecuando de acuerdo al ritmo de vida de la población salvadoreña, y al progreso y cambio que se ha generado desde entonces. El MINSAL debe, en el área médica, garantizar a los habitantes de la República de El Salvador la cobertura de servicios oportunos e integrales, con equidad, calidad y calidez; en corresponsabilidad con la comunidad, incluyendo todos los sectores y actores sociales, para contribuir a lograr una mejor calidad de vida. Por lo tanto, hoy en día, debe apoyar muchas de sus funciones en Sistemas Informáticos que le provean información rápida y oportuna para poder asegurar dicha calidad de vida con la que está comprometida.

Durante el desarrollo del proyecto de Trabajo de Graduación titulado "*Desarrollo e Integración de funciones estadísticas para el análisis de datos en el Sistema Integrado de Indicadores Gerenciales del Ministerio de Salud*" se ha buscado una solución que satisfaga las necesidades surgidas dentro de la institución en el área de análisis de datos, para lo cual se ha hecho uso de una metodología basada en el ciclo de vida de desarrollo de proyectos.

Este documento ha sido elaborado para registrar y describir las partes esenciales de todo el proceso de desarrollo del proyecto, destacando los elementos más importantes y relevantes durante todas y cada una de las etapas del proyecto.

A continuación se presenta una breve descripción de cada uno de los capítulos que conforman este documento para presentar al lector una antesala de su contenido.

## **Capítulo 1: Investigación Preliminar**

Se presenta una investigación previa de carácter bibliográfico y de campo que permite la creación de una base de conocimiento acerca del eje central del proyecto y los temas relacionados con el mismo.

## **Capítulo 2: Situación Actual**

Se presenta la situación actual en cuanto al Sistema Integrado de Indicadores Gerenciales (SIIG/eTAB) que ya se encuentra funcionando, se identifica qué se tiene actualmente para poder dejar claro cuál fue el punto de partida del proyecto.

Se hace uso del enfoque de sistemas como técnica para presentar un análisis del sistema actual, determinando sus entradas, procesos, salidas, medio ambiente y controles que actualmente se utilizan y se verán afectados durante el desarrollo del proyecto.

## **Capítulo 3: Análisis del Problema**

Se presenta con más detalle el problema al que se le busca dar solución mediante la realización de este trabajo, haciendo uso de la técnica de la caja negra se busca presentar mediante dicho diagrama el estado actual y el estado futuro que se pretende lograr al haber finalizado este proyecto.

También se hace uso de un enfoque de sistemas para presentar el estado futuro que se pretende lograr una vez completado el desarrollo de este proyecto.

## **Capítulo 4: Análisis de Requerimientos**

Se definen los requerimientos informáticos, operativos, no funcionales, técnicos y restricciones. El contenido de este capítulo representa una de las partes principales y de mayor importancia para el desarrollo del proyecto.

Se presentan los casos de uso del proyecto para permitir entender el proceso a seguir para cada uno de los pasos que conforman este proyecto. A la vez, también se presenta el procedimiento a seguir y que se desarrolló durante el desarrollo del proyecto para poder desarrollar de forma exitosa cada una de las funciones estadísticas.

## **Capítulo 5: Diseño**

Se presentan bosquejos de las principales pantallas tanto de entrada de datos como salida de resultados (o información) que representan una parte esencial del propósito final de este proyecto. Se desglosan una a una las funciones estadísticas y se permite tener una idea de los procesos involucrados en el desarrollo de las mismas.

## **Capítulo 6: Desarrollo del Proyecto e Implementación**

Se describen los estándares utilizados durante el desarrollo del proyecto. Se detalla el ámbito de programación utilizada, y se desglosa el código de programación empleado para desarrollar cada una de las funciones estadísticas desarrolladas a lo largo del proyecto.

Se presentan los pasos para llevar a cabo el proceso de integración del módulo de Análisis de Datos como parte funcional del Sistema Integrado de Indicadores Gerenciales (SIIG/eTAB), además se describe la documentación desarrollada para sustentar el desarrollo del proyecto y que permita hacer uso del nuevo módulo y realizar cualquier tipo de mantenimiento en el mismo.

También se ha desarrollado un plan de implementación para poder poner en funcionamiento el proyecto desarrollado de manera ordenada y efectiva en el menor tiempo posible.

## **Anexos**

Un conjunto de documentos cuyo contenido permite brindar información complementaria para expandir algún tema tratado dentro del contenido de este trabajo.

## OBJETIVOS

### GENERAL

- Desarrollar e Integrar Funciones para el Análisis de Datos haciendo uso del lenguaje *R-Statistics* en el Sistema Integrado de Indicadores Gerenciales (SIIG/eTAB), que sirva de apoyo en la toma de decisiones estratégicas.

### ESPECÍFICOS

- Analizar el Sistema Integrado de Indicadores Gerenciales SIIG/eTAB y los datos del mismo.
- Determinar los requerimientos informáticos, operativos y de desarrollo.
- Diseñar funciones estadísticas para el análisis de los datos que dan origen a los indicadores de salud haciendo uso del lenguaje *R-Statistics*.
- Construir e integrar al Sistema Actual nuevas funcionalidades en cuanto al análisis de los datos que dan origen a los Indicadores de Salud haciendo uso de herramientas de código abierto.
- Realizar las pruebas respectivas a las funciones estadísticas desarrolladas e integradas dentro del Sistema Actual de acuerdo a las necesidades de la Dirección de Tecnologías de Información y Comunicaciones (DTIC).
- Elaborar la documentación requerida por la DTIC del Ministerio de Salud: Manual de usuario y Manual técnico.
- Crear un plan de implementación que describa los pasos a seguir y las configuraciones necesarias para poner en marcha las nuevas funciones desarrolladas.

## **ALCANCES**

- El presente proyecto está orientado a ampliar las posibilidades que existen para el análisis de los datos que dan origen a los indicadores de salud definidos por un comité técnico de indicadores encargado de su definición y descripción.
- El almacén de datos del Ministerio de Salud que será utilizado por las funciones estadísticas para el análisis de datos recopila datos históricos de los 14 departamentos de El Salvador desde hace 9 años. Este almacén de datos se genera a través de un módulo encargado de proveer los datos necesarios y ya está actualmente integrado dentro del sistema actual con el que se integrará el Modulo de Análisis de Datos.
- Los resultados generados servirán de base y reforzarán los criterios a tomar en cuenta para la elaboración de una planificación.
- La finalidad del proyecto es apoyar la toma de decisiones en las áreas de prevención y planeación a través del desarrollo y aplicación de fórmulas estadísticas para el análisis de los datos.

## IMPORTANCIA

El Proyecto “Desarrollo e Integración de funciones estadísticas para el análisis de datos en el Sistema Integrado de Indicadores Gerenciales del Ministerio de Salud” es importante por las siguientes razones:

- Partiendo que actualmente no se cuenta con una herramienta de análisis de datos que utilice métodos estadísticos, su desarrollo representa un importante progreso en el área de salud, por los resultados que proporcionará y por el impacto que generará para apoyar la toma de decisiones.
- Habrá una base estadística en la cual apoyarse para la toma de decisiones y construcción de políticas de salud.
- Se contará con una base de conocimiento para la toma de decisiones, pues la herramienta proporcionara resultados en forma de tablas y graficas que permitirán una mejor visión al momento de tomar una decisión.
- Las tablas y graficas generadas a partir de los datos de los indicadores de salud estarán siempre disponibles cuando el usuario los necesite, ya que se generaran a partir de los almacenes de datos que posee el Ministerio de Salud.
- Acorde al programa “Construyendo la Esperanza” se beneficiará desde una perspectiva especifica como son consejos de gestión municipales, hasta una perspectiva general en el consejo nacional que permita definir estrategias centralizadas o parcializadas de salud.

Por las razones anteriormente mencionadas, es importante que se desarrolle el presente proyecto, pues será una importante herramienta de apoyo para los usuarios del Ministerio de Salud que se dedican a tomar decisiones en todos los niveles de organización de los consejos de gestión.

## JUSTIFICACIÓN

En el año 2009, se creó un plan llamado “Construyendo la Esperanza<sup>1</sup>” en el cual se recomendaba crear los consejos de gestión a niveles nacional, departamental y municipal. Dichos consejos de gestión contarían con la participación de profesionales de salud, sindicatos, cooperativas, ONGs presentes, entre otros, con el objetivo de contribuir al desarrollo humano integral de la comunidad por medio de la elaboración de políticas y planes de salud.

A medida que estos consejos de gestión empezaron a funcionar, se encontraron con la necesidad de contar con una Sistema Informático que les ayudara en la toma de decisiones a nivel estratégico, ya sea en su municipio, departamento o a nivel nacional. Tomando en cuenta dicha necesidad se creó el Sistema Integrado de Indicadores Gerenciales (SIIG/eTAB) con apoyo de la OPS<sup>2</sup>, en el cual analizan los datos de los diferentes sistemas transaccionales que registran los datos de los servicios de salud y, a partir de estos datos, se generan los indicadores de salud.

La cantidad de registros que se manejan a través del SIIG/eTAB para la generación de sus indicadores ronda aproximadamente por los 25, 000,000.

Como ejemplos de las cantidades de datos a tratar se pueden mencionar dos de los principales Sistemas Informáticos que se utilizan dentro de la red del Sistema de Salud Pública:

- Sistema de Morbi-Mortalidad (14, 000,00 de registros al año)
- Sistema de Vigilancia Epidemiológica (800,000 registros al año)

Con lo anterior queda en evidencia que el volumen de datos accesibles para el Sistema Informático actual es enorme y la tarea de ofrecer un análisis de los mismos se vuelve imposible sin el uso de herramientas de análisis como son las funciones estadísticas, ya que al aplicarse en conjunto a través de un Sistema Informático se permite manipular los datos de manera oportuna generando resultados satisfactorios.

Además, como sucede con todo Sistema Informático, con el pasar de los tiempos surgen nuevas oportunidades de mejora que permiten obtener nuevos resultados y aprovechar un poco más las bondades ya ofrecidas por un Sistema Informático; es así como el MINSAL, a través de la Dirección de Tecnologías de Información y Comunicaciones (DTIC), mostró mucho interés en

---

<sup>1</sup>“Construyendo la Esperanza” Estrategias y Recomendaciones en Salud 2009-2014

<sup>2</sup>Organización Panamericana de la Salud

el desarrollo y aplicación de funciones estadísticas para el análisis de datos dentro de un módulo integrado al SIIG/eTAB.

Es bajo este contexto que surgió la oportunidad de desarrollar el Modulo de Análisis de Datos aplicando funciones estadísticas sobre los datos que dan origen a los indicadores de salud e incorporarlo dentro del Sistema Informático de Indicadores Gerenciales de Salud, y así, cubrir una necesidad del MINSAL e indirectamente, contribuir a mejorar la salud de la población salvadoreña.

---

---

# CAPÍTULO 1: INVESTIGACIÓN PRELIMINAR

## **SINOPSIS**

En este capítulo se presenta una investigación previa de carácter bibliográfico que permite la creación de una base de conocimiento acerca del eje central del proyecto y los temas relacionados con el mismo.

---

---

## 1.1 EL ALMACÉN DE DATOS DEL MINISTERIO DE SALUD

El Ministerio de Salud de El Salvador cuenta con distintos Sistemas Informáticos que son utilizados por sus distintas dependencias para llevar un control de las actividades desarrolladas como parte de los servicios que ofrecen. Estos Sistemas Informáticos registran sus datos en bases de datos operativas propias o compartidas que tienen como destino final formar parte de una base de datos central.

La base de datos central se convierte en un almacén de datos, llamado también *data warehouse*, después de aplicársele un proceso de Extracción, Transformación y Carga (ETL) con el objetivo de filtrar, seleccionar y normalizar los datos de acuerdo al uso y tratamiento que se les desea dar.

Para el caso de los Indicadores de Salud, se cuenta con un Sistema Informático de Indicadores Gerenciales de Salud (SIIG), también conocido como Tablero Electrónico de Indicadores (eTAB) y el cual es conocido por el acrónimo SIIG/eTAB o simplemente eTAB. Y es este Sistema Informático que hace uso del almacén de datos destinado a los Indicadores de Salud.

### 1.1.1 ELEMENTOS QUE CONSTITUYEN LA ARQUITECTURA DE UN ALMACÉN DE DATOS<sup>3</sup>

#### A. Base de datos operacional y base de datos externa

Las organizaciones adquieren datos de las bases de datos externas a la propia organización, dichas bases de datos externas incluyen datos demográficos, económicos, datos sobre la competencia, etc.

Mediante el proceso de *data warehousing* se extrae la información que está en la bases de datos operacionales y se mezcla con otras fuentes de datos. Con esto, enriquecemos la información que puede generarse a partir de estos datos almacenados.

---

<sup>3</sup> *Arquitectura Almacén de Datos*. Visitado 10/06/2015.< <http://anabuigues.com/2010/03/05/arquitectura-de-un-data-warehouse/>>

## **B. Nivel de acceso a la información**

Es la capa con la que trata el usuario final. La información generada, y en ocasiones almacenada, a partir de los datos se convierte en información fácil y transparente para las herramientas que utilizan los usuarios. Se obtienen informes, gráficos, diagramas, etc.

## **C. Nivel de acceso a los datos**

Comunica el nivel de acceso a la información con el nivel operacional, es el responsable de la interfaz entre las herramientas de acceso a la información y las bases de datos.

La clave de este nivel está en proveer al usuario de un acceso universal a los datos, es decir, que los usuarios sin tener en cuenta la ubicación de los datos o la herramienta de acceso a la información, deberían ser capaces de acceder a cualquier dato del *data warehouse* que les fuera necesario para realizar su trabajo.

## **D. Nivel de directorio de datos (metadatos)**

Para proveer de un acceso universal, es absolutamente necesario mantener alguna clase de directorio de datos o repositorio de información de metadato que ayude a mantener un control sobre los datos. El metadato aporta información sobre los datos de la organización, de dónde proviene, qué formato tenía, cuál era su significado y si se trata de un agregado, cómo se ha calculado éste.

Para mantener un almacén completamente funcional, es necesario disponer de una amplia variedad de metadatos, información sobre las vistas de datos para los usuarios finales y sobre las bases de datos operacionales.

## **E. Nivel de gestión de procesos**

Este nivel tiene que ver con la planificación de las tareas que se deben realizar, no sólo para construir, sino también para mantener el *data warehouse* y la información del directorio de datos. Es el controlador de alto nivel de los procesos que se han de llevar a cabo para que el *data warehouse* permanezca actualizado.

## **F. Nivel de mensaje de la aplicación**

Este nivel es el encargado del transporte de la información a lo largo del entorno, se puede pensar en él como un *middleware*.

### **G. Nivel *Data Warehouse* (físico)**

Es el núcleo del sistema, el repositorio central de información donde los datos actuales usados principalmente con fines informacionales residen. En el *data warehouse* físico se almacenan copias de los datos operacionales y/o externos, en una estructura que optimiza su acceso para la consulta y que es muy flexible.

### **H. Nivel de organización de datos**

Incluye todos los procesos necesarios para seleccionar, editar, resumir, combinar y cargar en el *data warehouse* y en la capa de acceso a la información los datos operacionales y/o externos.

## **1.1.2 ESTRUCTURA DE UN ALMACÉN DE DATOS**

La estructura de un almacén de datos (*data warehouse*) se caracteriza por los diferentes niveles de esquematización y detalle de los datos que se encuentran en él.

### **A. Detalle de los datos actuales**

Reflejan los acontecimientos más recientes, las últimas informaciones generadas por los sistemas de producción de la organización. El nivel de detalle no tiene por qué ser el mismo que el de los sistemas de producción, ya que los datos pueden ser fruto de alguna agregación o de una simplificación de los datos originales.

Una agregación es una partición horizontal de una relación según los valores de los atributos, seguida de una agrupación mediante una función de cálculo (suma, media, producto, etc.).

### **B. Detalle de datos antiguos**

Están almacenados en un nivel de detalle consistente con los datos detallados actuales, esto significa que si los datos actuales hacen referencia a ventas diarias en el año actual, los datos históricos contienen las ventas de años anteriores en el nivel de detalle de día también.

### **C. Datos resumidos**

Son datos obtenidos como resultado de un proceso de síntesis de los datos actuales. Lo que se tiene entonces son datos agregados o resumidos. Por ejemplo, se entiende mejor la evolución de las ventas si se presenta resumida por semanas que de manera diaria.

#### **D. Metadatos**

Ofrecen información descriptiva sobre el contexto, la calidad, la condición y las características de los datos. El metadato se sitúa en una dimensión diferente a la de los otros datos en el *data warehouse*, debido a que su contenido no es tomado directamente desde el ambiente operacional.

### 1.1.3 ARQUITECTURA DEL ALMACÉN DE DATOS DEL MINISTERIO DE SALUD

La siguiente Fig.1.1 representa la arquitectura del almacén de datos del Ministerio de Salud:

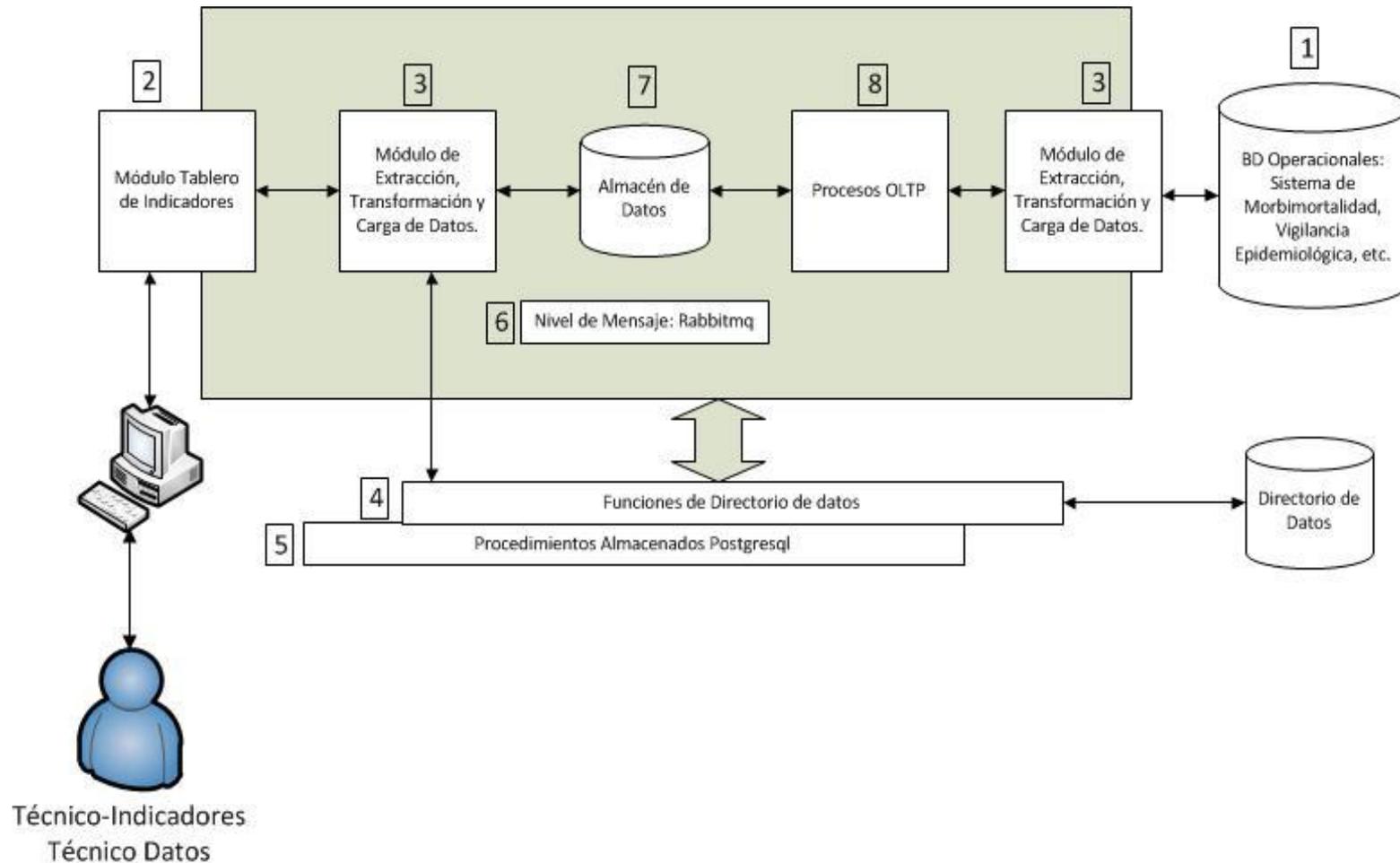


Fig.1.1-Arquitectura Almacén de Datos MINSAL

## 1.1.4 DESCRIPCIÓN DE LA ARQUITECTURA DEL ALMACÉN DE DATOS DEL MINISTERIO DE SALUD

Elementos de Arquitectura Data Warehouse	Descripción de elementos Data Warehouse MINSAL
1. Base de datos operacional y externa	Bases de datos operativas del MINSAL, por ejemplo vigilancia epidemiológica y morbimortalidad, además de hojas de cálculo y archivos en formato CSV.
2. Nivel de acceso a la información	Tablero de indicadores: Este módulo permite el acceso a la información utilizando visualizaciones graficas de los datos.
3. Nivel de acceso a los datos	El módulo ETL permite la extracción, transformación y carga de datos, estos son agregados y almacenados en una base de datos relacional (OLTP). Los datos están organizados por catálogos de referencia e indicadores medibles.
4. Nivel de directorio de datos	El módulo ETL permite a los usuarios administrar los indicadores y sus catálogos, así como todos los atributos de las tablas, es decir permite acceso a la metadata.
5. Nivel de gestión de procesos	Procedimientos almacenados en PostgreSQL permiten la actualización del almacén de datos.
6. Nivel de mensaje de la aplicación	RabbitMQ es un <i>middleware</i> que desempeña la función de mensajería dentro del almacén de datos.
7. Nivel <i>Data Warehouse</i>	Por la naturaleza de los datos se utiliza el modelo de base de datos sin esquema genérico EAV. Las tablas EAV y demás tablas auxiliares son parte del OLTP. Esto facilita el manejo de datos de cualquier indicador sin importar sus propiedades.
8. Nivel de organización de datos	Los procesos OLTP y los procesos que permiten la construcción de los cubos OLAP forman parte de este nivel, esto se realiza mediante procedimientos almacenados en PostgreSQL.

Tabla 1.1. Descripción de Arquitectura del Almacén de Datos del MINSAL

### 1.1.5 ESTRUCTURA DEL ALMACÉN DE DATOS DEL MINISTERIO DE SALUD

<b>Elementos de estructura</b>	<b>Descripción de elementos almacén de datos MINSAL</b>
Detalle de los datos actuales	Mediante la ejecución de procesos, el almacén de datos es periódicamente actualizado, ya que estos datos influyen en el resultado de los indicadores.
Detalle de datos antiguos	Las tablas almacenan datos antiguos sobre indicadores de salud, esto se realiza ejecutando procesos de extracción, transformación y carga de datos.
Datos resumidos	Se obtienen datos resumidos mediante la construcción de cubos OLAP y gráficas.
Metadatos	El sistema actual permite el acceso a la metadata mediante el módulo ETL.

Tabla 1.2. Elementos Estructura Almacén de Datos

## 1.2 LOS PROCESOS EN ALMACENES DE DATOS<sup>4</sup>

Para poder generar el almacén de datos que se utiliza dentro del Sistema Informático de Indicadores Gerenciales de Salud (SIIG/eTAB), se realiza un proceso sobre la base de datos central de los sistemas transaccionales del Ministerio de Salud. Este proceso recibe el nombre de ETL<sup>5</sup> y se describe a continuación.

Los procesos ETL son una parte de la integración de datos, debido a que es un elemento importante cuya función completa el resultado de todo el desarrollo de la cohesión de aplicaciones y sistemas.

La palabra ETL corresponde a las siglas en inglés de:

- Extraer: *Extract*.
- Transformar: *Transform*.
- Cargar: *Load*.

Los procesos ETL son cruciales en la integración de datos. Todo proceso ETL consta de estas tres fases: extracción, transformación y carga.

---

<sup>4</sup> Adaptado de la información contenida en el Blog de Powerdata

<sup>5</sup> Término proveniente del inglés de las palabras *Extract-Transform-Load* que significan Extraer, Transformar y Cargar y comúnmente se abrevian como ETL.

## 1.2.1 REPRESENTACIÓN DE PROCESO ETL EN ALMACÉN DE DATOS DEL MINISTERIO DE SALUD

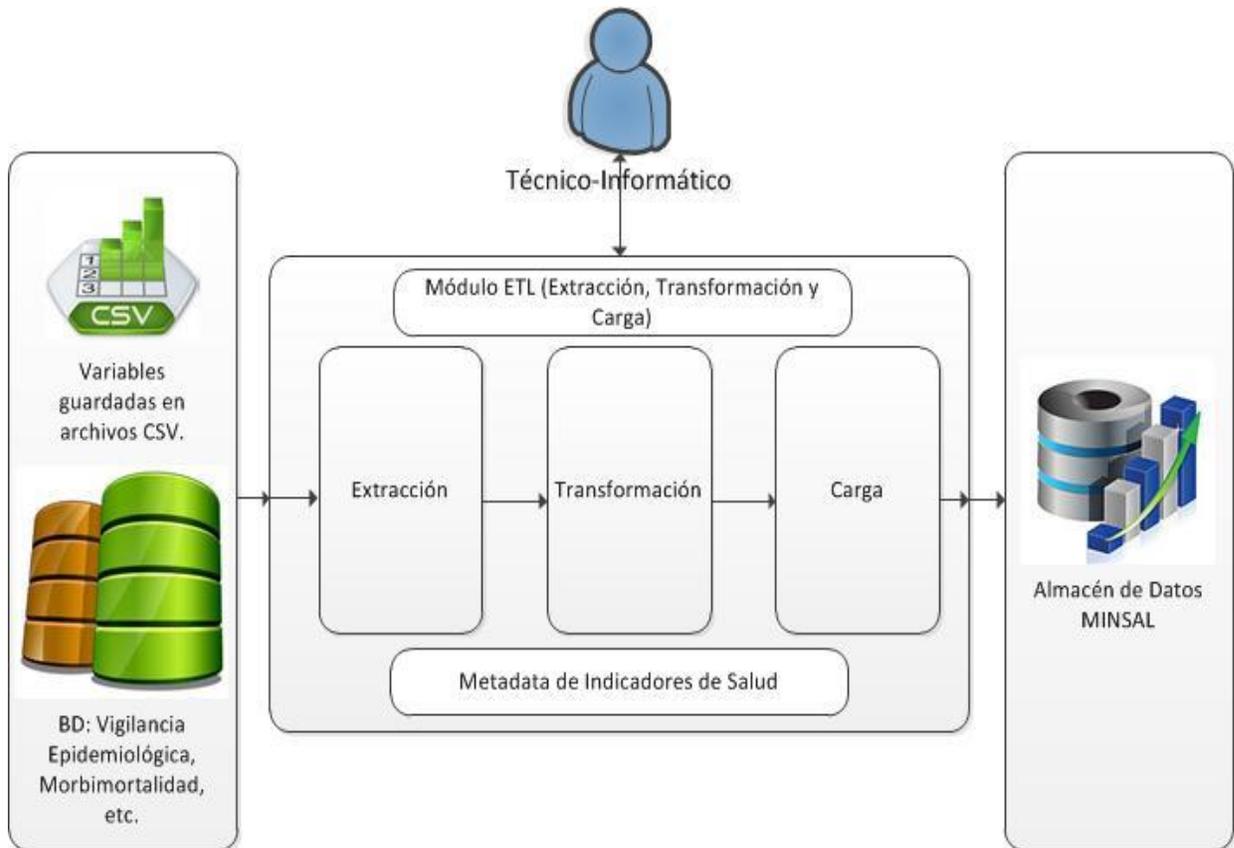


Fig.1.1-Arquitectura Almacén de Datos MINSAL

EL proceso ETL (Extracción, Transformación y Carga de Datos) es realizado por un módulo específico dentro del Sistema Integrado de Indicadores Gerenciales, a continuación se detallan las funciones de estos procesos.

### Fase de Extracción

Este proceso realiza las siguientes funciones:

- Extrae los datos desde los sistemas de origen.
- Analiza que los datos extraídos cumplan ciertas condiciones.

- Interpreta las condiciones que deben cumplirse para verificar que los datos extraídos cumplen la pauta o estructura que se esperaba. Si no es así, los datos deberían ser rechazados.
- Convierte los datos a un formato preparado para iniciar el proceso de transformación.

Además, uno de las consideraciones más importantes que se tienen en cuenta durante el proceso de extracción es exigir siempre que esta tarea cause un impacto mínimo en el sistema de origen.

Este requisito se basa en la práctica, ya que si los datos a extraer son muchos, el sistema de origen se podría ralentizar e incluso colapsar, provocando que no pudiera volver a ser utilizado con normalidad para su uso cotidiano.

### Fase de Transformación

La fase de transformación de un proceso de ETL aplica una serie de reglas de negocio o funciones, sobre los datos extraídos para convertirlos en datos que serán cargados. Estas directrices pueden ser declarativas, pueden basarse en excepciones o restricciones, pero para potenciar su pragmatismo y eficacia, hay que asegurarse de que sean:

- Declarativas.
- Independientes.
- Claras.
- Inteligibles.
- Con una finalidad útil para el negocio.

### Proceso de Carga

En esta fase, los datos procedentes de la fase anterior (fase de transformación) son cargados en el sistema de destino.

Existen dos formas básicas de desarrollar el proceso de carga:

1. **Acumulación simple:** esta manera de cargar los datos consiste en realizar un resumen de todas las transacciones comprendidas en el período de tiempo seleccionado y transportar el resultado como una única transacción hacia el data warehouse, almacenando un valor calculado que consistirá típicamente en un sumatorio o un

promedio de la magnitud considerada. Es la forma más sencilla y común de llevar a cabo el proceso de carga.

2. **Rolling:** este proceso sería el más recomendable en los casos en que se busque mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna o varias de las dimensiones de la magnitud almacenada (por ejemplo, totales diarios, totales semanales, totales mensuales, etc.).

## 1.3 EL ANÁLISIS DE DATOS

### 1.3.1 ANÁLISIS DE DATOS<sup>6</sup>

El análisis de datos es el proceso de inspeccionar, limpiar y transformar datos con el objetivo de resaltar información útil, lo que sugiere conclusiones, y apoyo a la toma de decisiones. El análisis de datos tiene múltiples facetas y enfoques, que abarca diversas técnicas en una variedad de nombres, en diferentes negocios, la Ciencia, y los dominios de las Ciencias Sociales.

El Análisis de Datos (Data Analysis, o DA) es usado en varias industrias para permitir que las compañías y las organizaciones tomen mejores decisiones empresariales y también es usado en las ciencias para verificar o reprobando modelos o teorías existentes. El análisis de datos se distingue de la extracción de datos por su alcance, su propósito y su enfoque sobre el análisis. Los extractores de datos clasifican inmensos conjuntos de datos usando software sofisticado para identificar patrones no descubiertos y establecer relaciones escondidas. El análisis de datos se centra en la inferencia, el proceso de derivar una conclusión basándose solamente en lo que conoce el investigador.

El término “análisis” ha sido usado por varios proveedores de software de inteligencia de negocios como una palabra de moda que describe varias funciones. El análisis de datos se usa para describirlo todo, desde el procesamiento analítico en línea (OLAP, por sus siglas en inglés) hasta el análisis CRM en centros de llamadas. Los bancos y las compañías de tarjetas de crédito, por ejemplo, analizan los retiros y los patrones de gasto para prevenir el fraude o robo de identidad. Las compañías de comercio electrónico (eCommerce) examinan el tráfico en el sitio web o los patrones de navegación para determinar qué clientes son más o menos propensos a comprar cierto producto o servicio, basándose en compras previas o patrones de visualización. El análisis de datos moderno normalmente usa tableros de información que se basan en flujos de datos en tiempo real. El llamado análisis en tiempo real implica análisis e informes dinámicos basados en los datos que han sido introducidos en un sistema un minuto antes del tiempo actual de uso.

---

<sup>6</sup> *Análisis de Datos*. Visitado el 01/04/2015. < <http://searchdatacenter.techtarget.com/es/definicion/Analisis-de-Datos>>

### 1.3.2 EL CONCEPTO DE ANÁLISIS ESTADÍSTICO DE DATOS<sup>7</sup>

El Análisis Estadístico de Datos se ha asociado de manera general con la investigación de corte experimental, o podemos decir que es característico de los enfoques positivistas. Este término se define de muchas maneras según se conciba desde una perspectiva más amplia o más restringida, y según se entienda también el proceso de investigación.

El concepto de Análisis Estadístico de Datos no se agota en las acepciones que se identifican con un conjunto de datos o enumeración de hechos, o con procedimientos de tipo descriptivo destinados a recoger, organizar y presentar la información relativa a un conjunto de casos. De esta manera, el Análisis Estadístico de Datos ha dejado de ser únicamente la ciencia de recopilar datos y, tras fusionarse con la corriente de estudios sobre el cálculo de probabilidades, se ha constituido en una rama de la matemática aplicada, entendiendo ésta como el uso de principios y modelos matemáticos en diversos ámbitos de la ciencia o la técnica.

Dentro del Análisis Estadístico de Datos encontramos dos vertientes:

1. Análisis Estadístico de Datos (propia mente matemático), que supone el estudio de los fenómenos estadísticos utilizando los métodos matemáticos y proporciona conocimiento acerca de las técnicas que integran los métodos estadísticos.
2. Análisis Estadístico de Datos Aplicado; este carácter aplicado ha estado presente desde los inicios de esta ciencia, sobre todo en cuanto a su conexión con el estudio y resolución de problemas prácticos con datos reales. Todo ello ha estimulado la innovación de nuevos métodos y procedimientos, y el avance de análisis estadísticos.

Las características propias que definen el concepto de Análisis Estadístico de Datos son:

- Posee un carácter teórico y aplicado.
- Realiza un estudio de un conjunto de datos.
- Le otorga carácter cuantitativo a los datos.
- Reducción de la información.
- Generalización a colectivos más amplios.

---

<sup>7</sup> *El concepto de análisis estadístico de datos.* Visitado el 01/04/2015. < [http://ocwus.us.es/metodos-de-investigacion-y-diagnostico-en-educacion/analisis-de-datos-en-la-investigacion-educativa/Bloque\\_1/page\\_03.htm](http://ocwus.us.es/metodos-de-investigacion-y-diagnostico-en-educacion/analisis-de-datos-en-la-investigacion-educativa/Bloque_1/page_03.htm) >

Por tanto, como ya se mencionó, el Análisis Estadístico de Datos no se limita sólo a su tratamiento sino que se extiende a tareas previas y posteriores a esta fase. También puede ocuparse de la recogida de datos (referido a las técnicas y métodos de muestreo y a la evaluación de la calidad de los instrumentos que se diseñan para la recogida de datos) y la interpretación de los resultados (afirmaciones que se realizan como consecuencia de la aplicación de métodos estadísticos: descripción, reducción, generalización).

Así, definimos el Análisis Estadístico de Datos como *un conjunto de métodos, técnicas y procedimientos para el manejo de datos, su ordenación, presentación, descripción, análisis e interpretación, que contribuyen al estudio científico de los problemas planteados en el ámbito en estudio y a la adquisición de conocimiento sobre una realidad determinada, a la toma de decisiones y a la mejora de la práctica desarrollada* (Gil Flores, 1996: 43).

El procedimiento de Análisis suele comenzar con una depuración de los datos para tratar de eliminar o corregir los posibles errores que se hayan cometido en la fase de registro o codificación de la información recogida. Una segunda fase es la descripción de las variables que entrarán a formar parte del estudio, teniendo en cuenta el enfoque del que partimos (univariante, multivariante o aplicación de técnicas inferenciales).

No obstante, existen vías diferentes al Análisis Estadístico de Datos. En este sentido, entre los procedimientos de Análisis de datos, entendidos como técnicas para extraer información de los datos e interpretar su significado, cabría hacer una distinción entre:

- *Análisis Cualitativo*, que expresa, ordena, describe, interpreta los datos mediante conceptos, razonamientos y palabras, y
- *Análisis Cuantitativo*, en el que se recurre a conceptos y razonamientos que se apoyan en números y estructuras matemáticas (Yela, 1994).

## 1.4 LA INFORMÁTICA EN EL ANÁLISIS ESTADÍSTICO DE DATOS<sup>8</sup>

De todas las funciones que la computadora puede cubrir en la investigación, es quizá la del Análisis de Datos la que con mayor claridad puede haberse beneficiado de este tipo de herramientas. La computadora ha venido a revolucionar este campo, permitiendo la aplicación de complejos procedimientos de análisis y, de alguna manera, simplificándolos.

Si el impacto de la informática en el campo de la investigación ha sido considerable, en el caso del Análisis de Datos esta afirmación cobra pleno sentido. El Análisis de Datos "ha recibido un impulso revolucionario con la generalización del uso de los computadores" (Bisquerra, 1989:230), hasta tal punto que hoy es difícil concebir este proceso desligado de la utilización de tales máquinas. Algunas de las ventajas y posibilidades que lleva consigo el uso de la informática en el Análisis Estadístico de Datos son:

- La computadora ha permitido un considerable ahorro de tiempo y esfuerzo. Los cálculos manuales que costaban al investigador largas horas de trabajo, incluso utilizando la calculadora, la máquina los realiza en pocos segundos.
- La computadora posibilita una mayor exactitud en los cálculos. Es evidente que el cálculo manual, además de ser lento conlleva aproximaciones o redondeos, sobre todo cuando se trabaja con números decimales y se requieren cálculos encadenados, que pueden llegar a suponer un considerable error en los resultados finales obtenidos.
- La computadora ha abierto la posibilidad de manejar grandes cantidades de datos, de trabajar con muestras mayores y de incluir más variables, haciendo que el análisis de grandes bancos de datos o la aplicación de complejas técnicas multivariantes a grandes muestras no presenten tareas imposibles en la práctica.
- Al liberarnos de tiempo en la realización de cálculos el investigador puede centrarse en otros momentos o tareas conceptuales más relevantes del proceso de investigación como la toma de decisiones respecto al proceso estadístico a seguir o a la técnica concreta que se deberá emplear, mayor atención al análisis crítico del proceso y a la interpretación de resultados.

---

<sup>8</sup> *La Informática en el Análisis Estadístico de Datos*. Visitado el 01/04/2015. < [http://ocwus.us.es/metodos-de-investigacion-y-diagnostico-en-educacion/analisis-de-datos-en-la-investigacion-educativa/Bloque\\_1/page\\_10.htm](http://ocwus.us.es/metodos-de-investigacion-y-diagnostico-en-educacion/analisis-de-datos-en-la-investigacion-educativa/Bloque_1/page_10.htm) >

- También nos ofrecen la posibilidad de realizar cálculos para la comparación de resultados utilizando métodos diferentes, aspecto que no se realizaba por considerar un esfuerzo adicional para el investigador, además de la dificultad que entrañaba un análisis.
- No obstante, el desarrollo de la Informática no sólo ha supuesto beneficios para el Análisis Estadístico de Datos sino también a la mejora y desarrollo de técnicas estadísticas más complejas e innovadoras.

Teniendo en cuenta las ventajas que representa el uso del ordenador y el papel que juega en el Análisis Estadístico de Datos, hoy día resulta inconcebible la realización de este tipo de análisis en la investigación sin el apoyo de la informática.

## 1.5 PRINCIPALES FORMULAS ESTADÍSTICAS PARA EL ANÁLISIS DE DATOS

### CUADRO RESUMEN DE LAS PRINCIPALES FUNCIONES ESTADÍSTICAS

CATEGORÍA	PROPÓSITO GENERAL	FUNCIÓN ESTADÍSTICA	PROPÓSITO ESPECIFICO
Medidas de Posición	Facilitan información sobre la serie de datos que estamos analizando.	Cuartiles	Dividir una distribución de datos en 4 partes iguales
		Deciles	Dividir una distribución de datos en 10 partes iguales
		Percentiles	Dividir una distribución de datos en 100 partes iguales
Medidas de tendencia central	Permiten identificar y ubicar los valores más representativos de los datos, de acuerdo a la manera como se tienden a concentrar.	Media	Se define como la suma de todos los valores observados, dividido por el número total de observaciones
		Mediana	Identifica el valor que se encuentra en el centro de los datos
		Moda	Indica el valor que más veces se repite dentro de los datos
Medidas de dispersión	Permiten reconocer que tanto se dispersan los datos alrededor del punto central.	Varianza	Permite identificar la diferencia promedio que hay entre cada uno de los valores respecto a su punto central
		Desviación típica	Permite determinar el promedio aritmético de fluctuación de los datos respecto a su punto central
		Coefficiente de variación	La relación entre el tamaño de la media y la variabilidad de la variable, se utiliza el coeficiente de variación
Medidas de distribución	Permiten identificar la forma en que se separan o aglomeran los valores de acuerdo a su representación gráfica.	Asimetría	Permite identificar si los datos se distribuyen de forma uniforme alrededor del punto central
		Curtosis o apuntamiento	Determina el grado de concentración que presentan los valores en la región central de la distribución
		Curva Normal	Permite modelar numerosos fenómenos naturales, sociales y psicológicos.
		Diagramas de cajas	Presentación visual que describe varias características importantes, al mismo tiempo, tales como la dispersión y simetría.

Tabla 1.3. Resumen Funciones Estadísticas.

CATEGORÍA	PROPÓSITO GENERAL	FUNCIÓN ESTADÍSTICA	PROPÓSITO ESPECIFICO
Estadística Descriptiva	Ayuda a presentar los datos de modo tal que sobresalga su estructura. Permite detectar tanto las características sobresalientes como las características inesperadas.	Correlación	Permite observar la fuerza y el sentido de una relación lineal y proporcionalidad entre dos variables estadísticas.
		Matriz de Correlaciones	Presenta numéricamente la correlación existente entre las diferentes variables expresada con valores entre -1 y 1.
		Mapa de correlaciones	Presenta gráficamente mediante círculos y colores la correlación que existe entre las diferentes variables.
		Circulo de correlaciones	Presenta mediante una serie de líneas rectas la correlación que existe entre las diferentes variables.
		Análisis de Conglomerados	Representar una realidad que no conseguimos visualizar, una realidad cuya representación original es multidimensional y es imposible que la podamos ver en estado puro.
		Centros de Gravedad	Representa el centro de la nube de puntos que forman parte de un conglomerado de valores.
Estadística Inferencial	Determina propiedades de una población estadística, a partir de una pequeña parte de la misma.	Estimación Puntual	Se usa un solo valor extraído de la muestra para estimar el parámetro desconocido de la población.
		Prueba de Hipótesis	Permite juzgar si una propiedad que se supone en una población estadística es compatible con lo observado en una muestra de dicha población

Tabla 1.4. Continuación de Resumen Funciones Estadísticas.

## 1.6 LOS INDICADORES ESTANDARIZADOS DE SALUD<sup>9</sup>

Los Ministros de Salud asistentes al Concejo Directivo de la Organización Panamericana de la Salud en el año 2000, aprobaron por unanimidad una resolución que recomienda a los países miembros a participar en la medición del desempeño de las Funciones Esenciales de Salud Pública, lo que obliga a rendir cuentas sobre el estado de salud de la población.

En ese sentido, el Ministerio de Salud se dio a la tarea de elaborar un instrumento guía que contuviera los indicadores seleccionados como priorizados del quehacer institucional, los cuales serán aplicados en los diferentes niveles de la estructura organizativa, en respuesta al cumplimiento de la primera Función Esencial de Salud Pública, a través de la dirección de planificación y la unidad de información en salud en coordinación con el comité de análisis central presento el documento: “Guía de indicadores priorizados y estandarizados en salud<sup>10</sup>”.

Un indicador es una medida que se puede usar para ayudar a describir una situación existente o para medir cambios o tendencias a lo largo de un periodo de tiempo, estos indicadores pueden ser cualitativos y cuantitativos.

Todo indicador está referido a:

- ✓ La población a partir de la cual se calculan.
- ✓ El periodo de tiempo que representan.
- ✓ El lugar geográfico del cual proviene la información (variables de persona, de tiempo y de lugar).

Los indicadores son importantes porque reúnen las siguientes características:

- Permiten apreciar cambios a través del tiempo.
- Facilitan ver de cerca los resultados de programas.
- Son instrumentos para evaluar y dar seguimiento a procesos.

---

<sup>9</sup> Para más información sobre los indicadores consultar Anexo 3

<sup>10</sup> “Guía de Indicadores Priorizados y Estandarizados en Salud” Mayo 2004

Los indicadores de salud pueden ser divididos en dos categorías: Indicadores epidemiológicos e Indicadores operacionales.

Los indicadores epidemiológicos se utilizan para estimar la magnitud, trascendencia y vulnerabilidad de los diferentes grupos de población ante una situación de salud determinada, en un espacio y tiempo definido. Así mismo, los indicadores epidemiológicos permiten conocer la incidencia y prevalencia de enfermedades sujetas a vigilancia, así como la mortalidad de la población.

A continuación se detalla un ejemplo de indicador epidemiológico:

<b>Nombre de Indicador:</b>	Tasa de incidencia de Dengue confirmado.
<b>Definición Conceptual:</b>	Casos confirmados de dengue en la población.
<b>Actividad a Evaluar:</b>	Presencia de Dengue confirmado.
<b>Fórmula:</b>	$\left( \frac{\text{Número total de casos nuevos de Dengue}}{\text{Población total}} \right) \times 100,000.$
<b>Interpretación:</b>	Por cada 100,000 habitantes existen “x” casos de Dengue confirmado.

Los indicadores operacionales miden el trabajo realizado, ya sea en función de la cantidad o de la calidad de ellos, además, miden la cantidad de actividades y procedimientos realizados.

Las medidas de frecuencia utilizadas en cualquier tipo de indicador son las siguientes:

- **Cifras absolutas:** Es la expresión simple de cantidad, da una idea de la magnitud o volumen real de un suceso.
- **Razón:** Es el cociente de dos variables, los valores del numerador y del denominador son independientes, ninguno está contenido en el otro.
- **Proporción:** Es el cociente de dos variables, el numerador está contenido en el denominador.
- **Porcentaje:** La gran mayoría de los indicadores son proporciones para facilitar su interpretación.
- **Tasa:** Es la magnitud del cambio de la variable que mide un evento por unidad de cambio de otra, en relación al tamaño de la población que se encuentra en riesgo de presentar el evento.

---

---

# CAPÍTULO 2: SITUACIÓN ACTUAL

## **SINOPSIS**

En este capítulo se presenta mediante el uso del enfoque de sistemas el contexto actual que presenta el Sistema de Indicadores Gerenciales (SIIG/eTAB) para poder presentar el contexto en el que se incorpora lo realizado con este proyecto.

## 2.1 SITUACIÓN ACTUAL

### 2.1.1 DESCRIPCIÓN

El Ministerio de Salud cuenta con el Sistema Integrado de Indicadores Gerenciales (SIIG/eTAB), hoy llamado unoTAB, el cual es un sistema informático para dar seguimiento y control a los indicadores de salud que trabaja sobre los datos obtenidos a partir de los sistemas estatales de información de salud.

Los indicadores de salud del SIIG/eTAB son definidos por el Comité Técnico de Indicadores para dar respuesta a organismos internacionales como la OPS sobre la salud de la población. Además este comité es multidisciplinario y es el encargado de definir de qué sistemas de información se tomarán los datos para ser procesados por el SIIG/eTAB.

Este sistema es un servicio Web disponible para que dependencias del sistema de salud suban sus datos y poder analizarlos de forma multidimensional, generar gráficas y reportes.

La aplicación cuenta con tres módulos:

- Módulo de Extracción, Transformación y Carga de Datos (ETL):  
Tiene como propósito efectuar la extracción, transformación y carga de datos (ETL) desde diferentes fuentes. Estos datos son agregados y almacenados en una base de datos relacional (OLTP).
- Módulo de Ficha Técnica del Indicador  
Se encarga de organizar los datos por catálogos de referencia e Indicadores medibles. Los usuarios del sistema pueden administrar estos indicadores y catálogos y todos sus atributos usando las herramientas que brinda el sistema.
- Módulo de Tablero de Indicadores:  
Permite realizar consultas en línea y los datos son agregados dentro de tablas optimizadas para el análisis en línea (OLAP, por sus siglas en inglés).

De forma general, la funcionalidad del sistema es la siguiente:

- Provee la capacidad de analizar datos de forma multidimensional para cualquier indicador.
- Cada indicador está construido con varios datos, los cuales están relacionados por una fórmula almacenada en el sistema.
- Es posible crear una tabla por cada grupo de datos, con la limitante, que es necesario conocer el dato antes de guardarlo, este comportamiento no es sostenible a futuro.
- La base de datos guarda los datos sin conocer de antemano sus características. Esto se logra usando un esquema de datos genérico EAV (entidad-atributo-valor). El manejo de esquemas EAV es implementado de diferentes formas para diferentes gestores de bases de datos.

Los resultados del sistema actual son utilizados por los diferentes consejos de gestión entre ellos están: El Consejo de Gestión de la Red Nacional, Consejo de Gestión de la Red Regional, Consejo de Gestión de la Red Departamental y El Consejo de Gestión de la Micro Red. Cada consejo puede ejecutar medidas y tomar decisiones de forma independiente, pero también deben ejecutar medidas a partir de decisiones tomadas a nivel de gabinete. Todos los consejos se reúnen para analizar diferentes problemáticas en cuanto a la salud de la población y se toman decisiones en común acuerdo.

Actualmente no existen procedimientos para el análisis de datos y no hay una unidad que se dedique a esta área, más bien, algunas dependencias del Ministerio de Salud hacen el análisis de los datos utilizando hojas de cálculo o herramientas estadísticas como SPSS<sup>11</sup>, PSPP<sup>12</sup>, e incluso de forma básica R-Statistics.

Los resultados actualmente obtenidos del análisis de datos de algunas dependencias como por ejemplo Vigilancia Epidemiológica no son del todo satisfactorios, es por esto que el Ministerio de Salud recibe cooperación de otras instituciones que realizan sus análisis de datos como la OPS y CDC de Atlanta. Si bien estos análisis son importantes, muchas veces no reflejan la realidad de la salud de la población.

---

<sup>11</sup> Programa estadístico informático muy usado en las ciencias sociales y las empresas de investigación de mercado.

<sup>12</sup> Aplicación de software libre para el análisis de datos

## 2.1.2 REPRESENTACIÓN MEDIANTE ENFOQUE DE SISTEMAS

### MEDIO AMBIENTE

LA ORGANIZACIÓN PANAMERICANA DE LA SALUD (OPS), BANCO INTERAMERICANO DE DESARROLLO (BID).

### FRONTERA

Red de Salud Pública de El Salvador

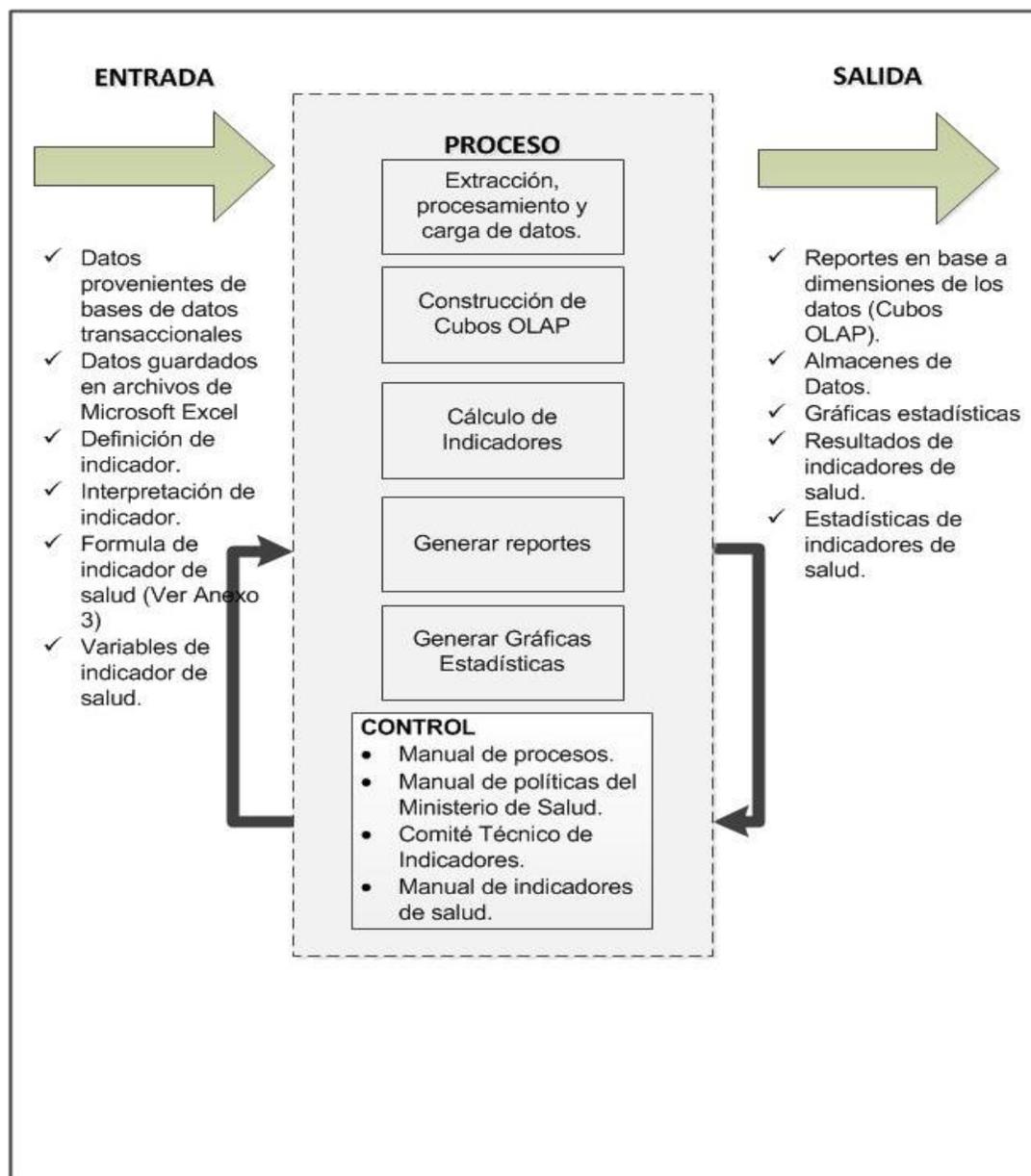


Fig.2.1 – Enfoque de Sistemas Situación Actual.

## 2.1.3 DESCRIPCIÓN DE LOS ELEMENTOS DEL ENFOQUE DE SISTEMAS

### MEDIO AMBIENTE

Es todo lo que está fuera del sistema y que se interrelaciona con dicho sistema.

- **La Organización Panamericana de la Salud (OPS):** Es el organismo especializado de salud del sistema interamericano, se encarga de evaluar el progreso en cuanto a salud se refiere, impulsa la creación de indicadores y hace recomendaciones en políticas de salud.
- **Banco Interamericano de Desarrollo (BID):**  
Es una organización financiera internacional con el propósito de financiar proyectos viables de desarrollo económico, social e institucional y promover la integración comercial regional en el área de América Latina y El Caribe.

### FRONTERA

La frontera de un sistema es el límite real o virtual de la extensión del sistema.

- **Red de Salud Pública de El Salvador:** La frontera de sistema está determinada por toda la Red de Salud Pública de El Salvador, dentro de esta red se encuentran los consejos de gestión nacional, consejos regionales, consejos departamentales, hospitales de primer nivel, etc.

### ENTRADAS

Es lo que el sistema recibe de su medio ambiente.

- **Datos provenientes de bases de datos transaccionales:** Son bases de datos de cada sistema informático en funcionamiento en El Ministerio de Salud, estas bases de datos alimentan al almacén de datos que se encuentra centralizado y estos datos son utilizados para analizarlos.
- **Datos guardados en archivos de Microsoft Excel:** Para el cálculo de indicadores de salud, a veces es necesario hacer uso de archivos de Excel, ya que esta información no se encuentra en las bases de datos transaccionales.
- **Definición del indicador:** Representa la definición conceptual de cada indicador de salud.

- **Interpretación del indicador:** Establece como se debe interpretar un indicador específico de acuerdo a la fórmula definida.
- **Fórmula del indicador de salud:** Son fórmulas<sup>13</sup> para el cálculo de características propias en el área de salud.
- **Variables de los indicadores de salud:** Son cada una de las variables que conforman a un indicador de salud específico.

## SALIDAS

Es lo que el sistema entrega a su medio ambiente.

- **Reportes en base a dimensiones de los datos (Cubos OLAP):** Es una herramienta para el análisis de los datos del indicador de forma dinámica donde el usuario puede seleccionar la combinación de datos que le resulte más efectiva para analizar la información disponible.
- **Almacenes de Datos:** Son colecciones de datos que están orientados a un determinado ámbito, en este caso es el ámbito de salud, están integrados, no volátiles y variables el tiempo. Todos los datos son obtenidos de las bases de datos transaccionales de todas las dependencias del Ministerio de Salud.
- **Graficas Estadísticas:** Los resultados de los indicadores<sup>14</sup> se pueden graficar para una mejor visualización y comprensión de los datos.
- **Resultados de Indicadores de salud:** Representan valores obtenidos en las fórmulas<sup>15</sup> de cada indicador de salud.
- **Estadísticas de Indicadores de salud:** Son impresas en pantalla o reportes donde se detallan indicadores de salud por región, por departamento, por municipio, anual, trimestral, mensual. El usuario determina cual indicador utilizar y que estadística necesita, variando espacio y tiempo.

## PROCESOS

Mecanismo que transforma las entradas en salidas dentro del sistema.

- **Extracción, procesamiento y carga de datos:** Permite la conexión a las bases de datos. En este proceso se registran los parámetros necesario para poder realizar la

---

<sup>13</sup> Para más información sobre las fórmulas de los indicadores véase Anexo 3

<sup>14</sup> Ídem

<sup>15</sup> Ídem

conexión y extracción de los datos necesarios que se utilizan para el cálculo del indicador respectivo, se trabaja sobre los datos obtenidos a partir de la información cuantificable válida, relevante y oportuna producidas por los diferentes sistemas de información transaccionales del Ministerio de Salud. La información disponible y accesible, permite una mejor coordinación y operación de la salud desde el MINSAL hacia otros organismos que requieran dicha información.

- **Generación de cubos OLAP:** Proceso por medio del cual se determinan los datos a utilizar de acuerdo a los indicadores que se requieran analizar, así como también se definen las dimensiones que tendrá el cubo OLAP.
- **Calculo de indicadores:** En base a fórmulas<sup>16</sup>ya establecidas, se realizan los cálculos respectivos para los indicadores de salud.
- **Generar reportes:** Mediante este proceso se generan cada uno de los reportes requeridos por los usuarios finales, que no son más que estadísticas de los indicadores de salud que varían en espacio y tiempo.
- **Generar graficas estadísticas:** Este proceso permite crear gráficos de cada indicador de salud, así como realizar consolidaciones.

## CONTROL

Es un mecanismo que opera dentro del sistema. Su función es monitorear las salidas del sistema con el fin de detectar toda desviación con respecto al objetivo del sistema; y emitir acciones correctivas cuando ocurren tales desviaciones.

- **Manual de procesos:** Es una forma de control, ya que establece como deben realizarse los procesos actuales que influyen directamente en el funcionamiento de los sistemas y herramientas de software.
- **Políticas del Ministerio de Salud:** Principios generales que el Ministerio debe cumplir, esto afecta directamente a los sistemas de información, ya que deben reajustarse si no se están cumpliendo los principios.
- **Comité Técnico de Indicadores:** Es un comité dentro del Ministerio de Salud que se encarga de crear nuevos indicadores y evalúan los resultados que se obtienen, así también verifican la correcta utilización y cálculo de dichos indicadores.

---

<sup>16</sup> Para más información sobre las fórmulas de los indicadores véase Anexo 3

- **Manual de Indicadores de Salud:** Contiene la definición de términos claves, actividades a evaluar, tipos de indicadores y sus estándares, periodicidad y construcción de los mismos, así como la interpretación y fuente de información respectiva.

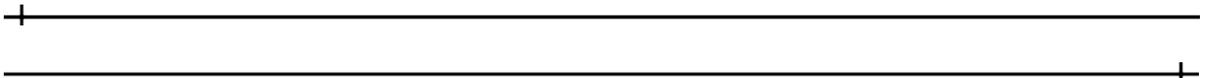
---

---

# **CAPÍTULO 3: ANÁLISIS DEL PROBLEMA**

## **SINOPSIS**

En este capítulo se presenta el problema que se busca solucionar mediante el desarrollo del este proyecto, se utiliza el diagrama de caja negra y el enfoque de sistemas para facilitar la presentación del análisis realizado.



### 3.1 PLANTEAMIENTO DEL PROBLEMA

La entidad encargada de la función informática dentro del Ministerio de Salud es la Dirección de Tecnología y Desarrollo Informático del Ministerio de Salud, la cual tiene como responsabilidades analizar y desarrollar sistemas informáticos, administrar las redes de comunicación y dar soporte técnico dentro del Ministerio de Salud.

Los principales Sistemas Informáticos que se utilizan para el apoyo de las operaciones del Ministerio y cuyos datos son de mucha relevancia para el análisis y toma de decisiones son:

- Sistema de Morbi-Mortalidad
- Sistema de Vigilancia Epidemiológica (VIGEPES)
- Sistema de Fichas Familiares
- Sistema de Encuestas del Ministerio de Salud

Además de los Sistemas Informáticos antes mencionados se tiene acceso a los datos pertenecientes a otros Sistemas Informáticos<sup>17</sup> (Ver Anexo 2) cuyo aporte es importante a la labor del Ministerio de Salud.

Actualmente el MINSAL cuenta con un Sistema Integrado de Indicadores de Salud cuya función principal es generar indicadores de salud al evaluar formulas determinadas haciendo uso de las variables asociadas a cada indicador. Dichos indicadores son presentados dentro del Sistema como parte de un tablero electrónico e interactivo que permite representar gráficamente cada indicador generado y alertar sobre cualquier comportamiento anómalo que pueda darse.

Para apoyar al MINSAL en la toma de decisiones estratégicas en medidas de salud y prevención se pretende el desarrollo de funciones estadísticas que se integrarán al Sistema Integrado de Indicadores de Salud para generar conocimiento a partir de los datos de los indicadores de Salud.

---

<sup>17</sup> Para más información sobre los sistemas consultar Anexo 2

Además, al utilizar métodos estadísticos sobre los indicadores de Salud generados y sus variables asociadas; se pueden detectar comportamientos similares entre los datos que permitan determinar estadísticamente asociaciones y dependencias entre las variables, conocidas como correlaciones. También es posible agrupar los datos de acuerdo a características en común como la región del centro de salud que ha recibido los casos, el nivel de educación del paciente, el estrato social, la edad, etc.

Todo lo anterior viene a mejorar los resultados que el Sistema Integrado de Indicadores de Salud presenta actualmente, ya que se limitaba a presentar gráficamente el comportamiento de acuerdo a la fórmula de cada indicador, más no tomaba en cuenta las características propias de cada variable asociada y de cada variable con las demás, algo que el análisis de datos sí es capaz de realizar.

Este nuevo conocimiento busca encontrar los puntos fuertes y débiles dentro del accionar del Ministerio de Salud, de acuerdo a cada indicador que puedan ser respaldados en los datos que se poseen, ya que por ejemplo se podrá detectar posibles causas de la ocurrencia de casos de dengue, rangos de edades con mayor ocurrencia de sarampión y posibles factores de su ocurrencia, entre otros. Este nivel de profundidad en el análisis de los datos actualmente no se tiene dentro del Sistema Integrado de Indicadores de Salud, pero se requiere, si se busca poder obtener nuevos puntos de atención en los cuales enfocarse para poder apoyar en la planificación de nuevos planes y estrategias de salud.

### 3.1.1 MÉTODO DE LA CAJA NEGRA

El concepto de caja negra se refiere a un sistema cuyo interior no puede ser descubierto, cuyos elementos internos son desconocidos y que sólo puede conocerse “por fuera”, a través de manipulaciones externas o de observación externa.<sup>18</sup>

El funcionamiento de ésta consiste en que la entrada que llega sea procesada dentro de la caja, y la salida que proporcione sea la transformación de dicha entrada.<sup>19</sup>

Para explicar la problemática que se intenta resolver con el desarrollo del proyecto de Desarrollo e Integración de Funciones Estadísticas para el Análisis de Datos, se aplicara el método de la caja negra para presentar, en forma general, el resultado esperado.

Se toma como estado A, los sistemas de información gerencial en funcionamiento actualmente, los cuales están enfocados en el análisis de los datos mediante cubos OLAP. Mientras que en el estado B, una vez aplicado el proceso se obtendrá información a través de la aplicación de técnicas estadísticas descriptivas. A continuación se presenta el planteamiento del problema utilizando el método de la caja negra:

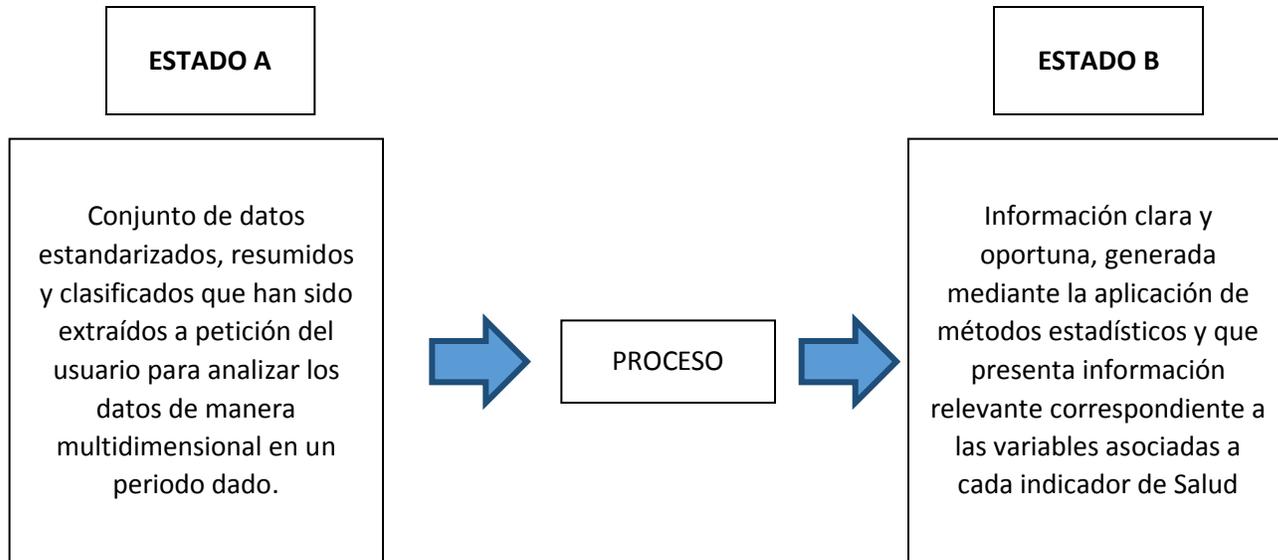


Fig.3.1 – Método de la Caja Negra para elaborar el Planteamiento del Problema.

<sup>18</sup> Caja Negra. Visitado el 03/04/2015. <<http://carolbr.wordpress.com/caja-negra/>>

<sup>19</sup>Ídem.

### 3.1.2 DETALLE DEL PLANTEAMIENTO DEL PROBLEMA

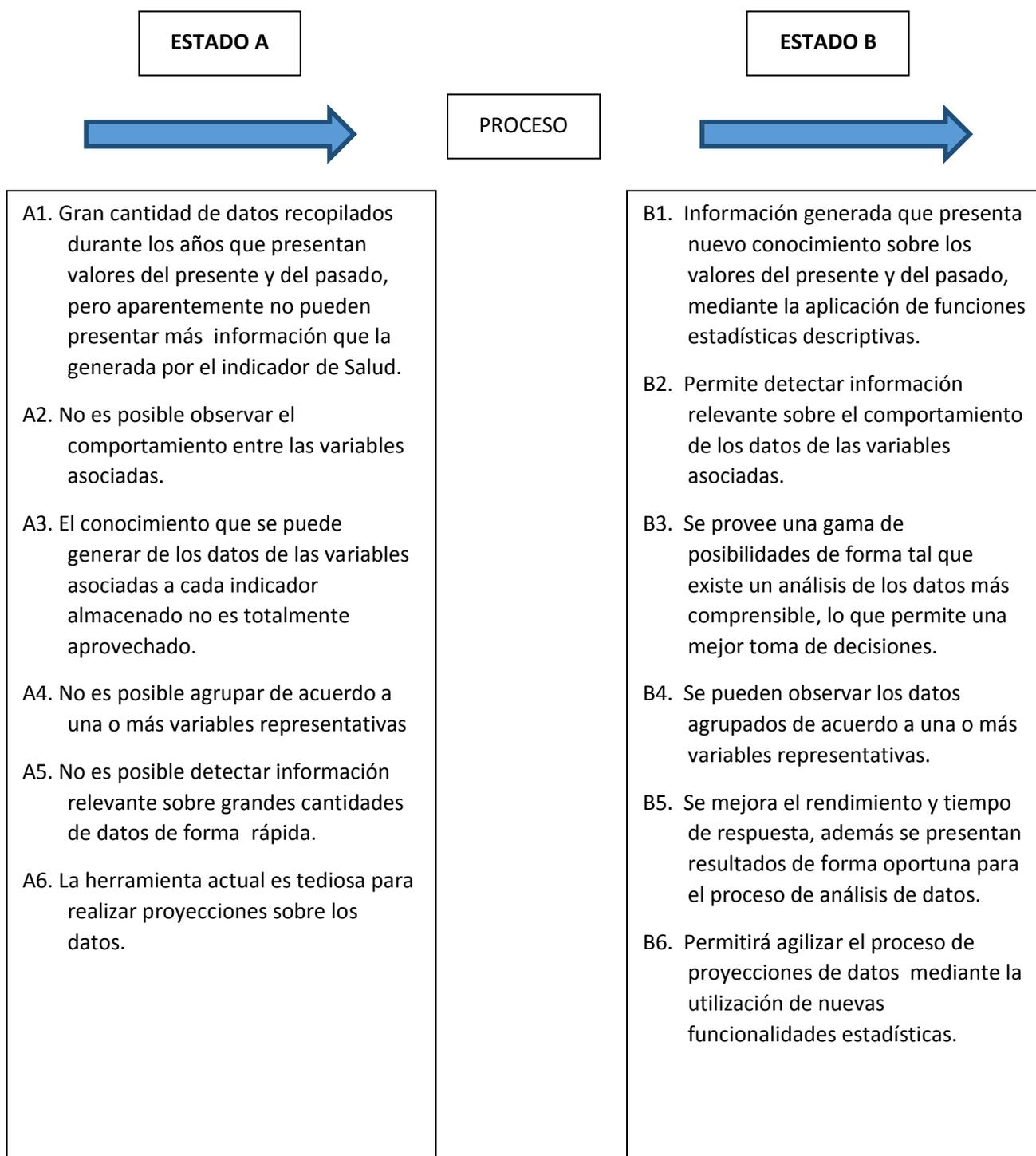


Fig. 3.2 – Detalle del Planteamiento del Problema.

## 3.2 FORMULACIÓN DEL PROBLEMA

Con la utilización del método de la caja negra, se puede formular el problema de la siguiente manera:

***¿En qué medida la utilización de funciones estadísticas para el análisis de datos de los indicadores de salud permitirá detectar las principales debilidades y/o fortalezas dentro de cada indicador de salud, así como también presentar información relevante que sirva de apoyo para la toma de decisiones a través de la generación de asociaciones, dependencias y agrupaciones de las variables que conforman a cada indicador?***

### 3.3 ENFOQUE DE SISTEMAS PROPUESTO

#### MEDIO AMBIENTE

LA ORGANIZACIÓN PANAMERICANA DE LA SALUD (OPS), BANCO INTERAMERICANO DE DESARROLLO (BID).

#### FRONTERA

Red de Salud Pública de El Salvador

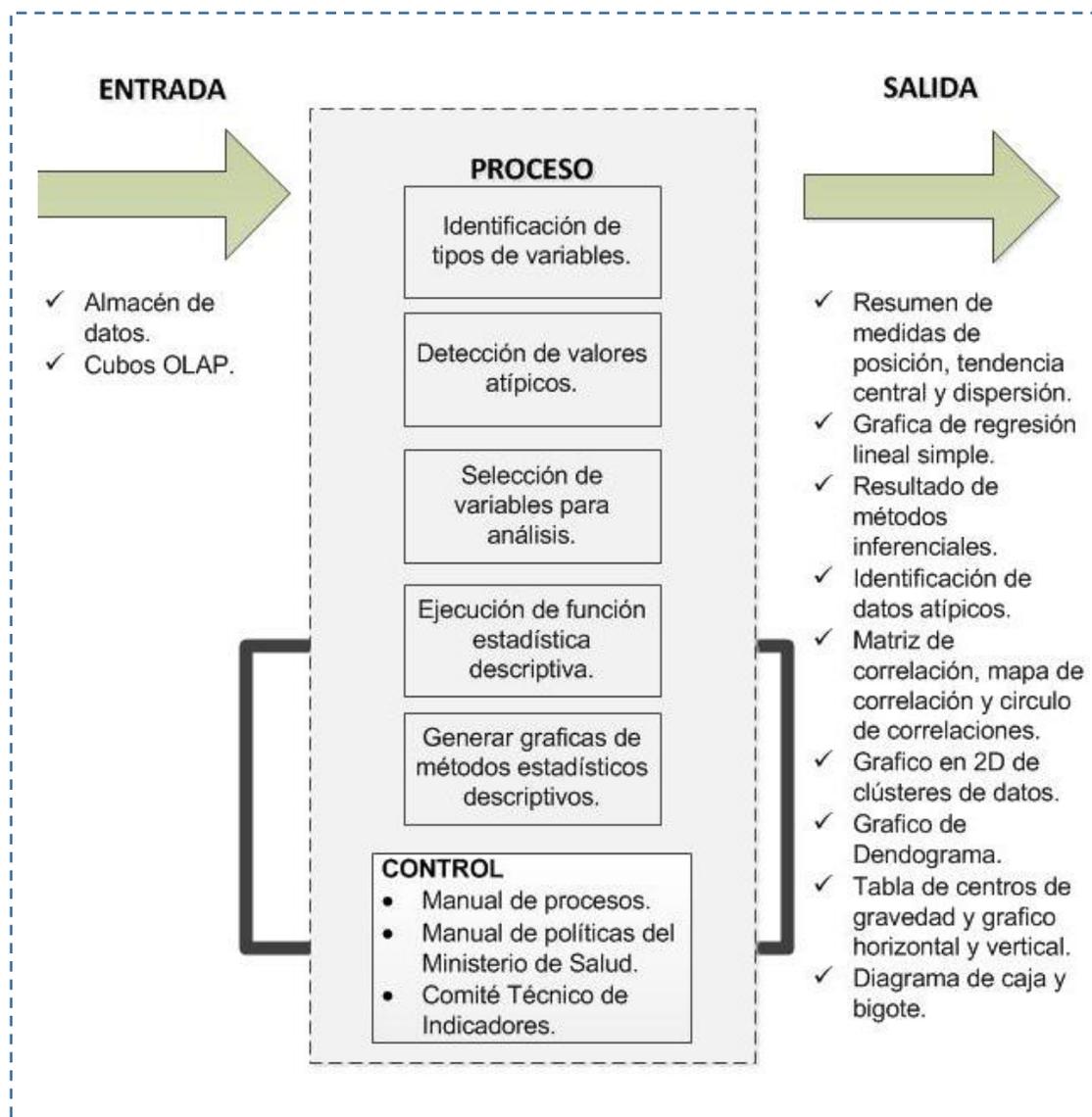


Fig.3.3-Enfoque de Sistemas Propuesto.

### 3.4 DESCRIPCIÓN DE ELEMENTOS ENFOQUE DE SISTEMAS PROPUESTO

#### MEDIO AMBIENTE

Es todo lo que esta fuera del sistema y que se interrelaciona con dicho sistema.

- **La Organización Panamericana de la Salud (OPS):** Es el organismo especializado de salud del sistema interamericano, se encarga de evaluar el progreso en cuanto a salud se refiere, impulsa la creación de indicadores, hace recomendaciones en políticas de salud.
- **Banco Interamericano de Desarrollo (BID):** Es una organización financiera internacional con el propósito de financiar proyectos viables de desarrollo económico, social e institucional y promover la integración comercial regional en el área de América Latina y El Caribe.

#### FRONTERA

La frontera de un sistema es el límite real o virtual de la extensión del sistema.

- **Red de Salud Pública de El Salvador:** La frontera de sistema está determinada por toda la Red de Salud Pública de El Salvador, dentro de esta red se encuentran los consejos de gestión nacional, consejos regionales, consejos departamentales, hospitales de primer nivel, etc.

#### ENTRADAS

Es lo que el sistema recibe de su medio ambiente.

- **Almacenes de Datos:** Son colecciones de datos que están orientados a un determinado ámbito, en este caso es el ámbito de salud, están integrados, no volátiles y variables el tiempo. Todos los datos son obtenidos de las bases de datos transaccionales de todas las dependencias del Ministerio de Salud.
- **Cubos OLAP:** Es una herramienta para el análisis de los datos del indicador de forma dinámica donde el usuario puede seleccionar la combinación de datos que le resulte más efectiva para analizar la información disponible.

## SALIDAS

Es lo que el sistema entrega a su medio ambiente.

- **Resumen de medidas de posición, tendencia central y dispersión:** Consiste en obtener un detalle de las principales medidas de posición (deciles, cuartiles, percentiles), de tendencia central (Media aritmética, Mediana) y de dispersión (Varianza, Desviación Típica) que puedan ser aplicadas a un indicador de salud que contenga variables cuantitativas.
- **Grafica de regresión lineal simple:** Grafico que representa la linealidad de los datos, es decir se utilizara para determinar si los datos de un indicador de salud cumplen con la función lineal estadística.
- **Resultado de métodos inferenciales:** Representa el valor de resultado obtenido al aplicar los métodos de estimación puntual y prueba de hipótesis.
- **Identificación de datos atípicos:** Consiste en identificar valores que se salen del rango común de los datos en cada indicador.
- **Matriz de correlación, mapa de correlación y circulo de correlaciones:** El cálculo de correlaciones busca de identificar que variables están asociadas entre si y cuales son independientes.
- **Gráfico en 2D de clústeres de datos:** Consiste en encontrar grupos de datos de similares características y se deben graficar en 2 dimensiones.
- **Gráfico de Dendograma:** El gráfico de dendograma es para identificar clústeres de datos, obteniendo una visualización detallada de los principales clústeres.
- **Tabla de Centros de gravedad y gráfico horizontal y vertical:** Permitirá identificar valores centrales en un clúster de datos.
- **Diagrama de caja y bigote:** Diagrama basado en cuartiles que permite la identificación de forma gráfica de valores atípicos.

## PROCESOS

Mecanismo que transforma las entradas en salidas dentro del sistema.

- **Identificación de tipos de variables:** Este proceso permitirá identificar el tipo de variables presentes (cualitativas o cuantitativas), antes de ejecutar una función estadística, con la finalidad de evitar una mala aplicación de las funciones estadísticas.

- **Detección de valores atípicos:** Se realizara una función que detecte valores atípicos en los datos a analizar, lo que permitirá mayor calidad en el resultado de las funciones.
- **Selección de variables para análisis:** Mediante este proceso el usuario podrá elegir que variables necesita para realizar el análisis respectivo.
- **Ejecución de función estadística descriptiva:** Consiste en el script de la función estadística, este es el principal proceso puesto que permitirá procesar los datos enviados como parámetros y ejecutara la función estadística correspondiente.
- **Generar graficas de métodos estadísticos descriptivos:** Este proceso es parte del script de la función estadística, permitirá la graficación de los resultados.

## CONTROL

Es un mecanismo que opera dentro del sistema. Su función es monitorear las salidas del sistema con el fin de detectar toda desviación con respecto al objetivo del sistema; y emitir acciones correctivas cuando ocurren tales desviaciones.

- **Manual de procesos:** Es una forma de control, ya que establece como deben realizarse los procesos actuales que influyen directamente en el funcionamiento de los sistemas y herramientas de software.
- **Políticas del Ministerio de Salud:** Principios generales que el Ministerio debe cumplir, esto afecta directamente a los sistemas de información, ya que deben reajustarse si no se están cumpliendo los principios.
- **Comité Técnico de Indicadores:** Es un comité dentro del Ministerio de Salud que se encarga de crear nuevos indicadores y evalúan los resultados que se obtienen, así también verifican la correcta utilización y cálculo de dichos indicadores.

---

---

# **CAPÍTULO 4: ANÁLISIS DE REQUERIMIENTOS**

## **SINOPSIS**

En este capítulo se presenta un listado de requerimientos que debe cumplir el proyecto, tanto en la parte operativa como en la parte técnica.

---

---

## **4.1 ANÁLISIS Y DETERMINACIÓN DE REQUERIMIENTOS**

Este proyecto busca incorporar nuevas características y funciones dentro de un Sistema Informático ya desarrollado, por lo que se trata de integrar los nuevos requerimientos sin dañar el funcionamiento actual.

Este apartado tiene por propósito especificar y clasificar de manera formal y oficial, los nuevos requerimientos funcionales que serán agregados al Sistema Integrado de Indicadores Gerenciales.

### **4.1.1 REQUERIMIENTOS FUNCIONALES**

1. El modulo contará con funciones estadísticas para los siguientes indicadores:
  - i. Casos de complicación de diabetes mellitus insulino dependiente.
  - ii. Casos de infección por VIH diagnosticados.
  - iii. Casos de defunciones de niños y niñas menores a un año.
  - iv. Casos de defunciones de niños menores a 5 años.
  - v. Casos de mortalidad fetal.
  - vi. Casos por muerte materna.
  - vii. Casos de muerte por cáncer de cuello uterino.
  - viii. Casos de muerte por cáncer de mama en mujeres.
  - ix. Casos de muerte por insuficiencia renal crónica.
  - x. Casos de hipertensión arterial.
  
2. Resumen de medidas de posición, de tendencia central y de dispersión para un indicador de salud determinado (varianza, media, cuartiles, etc.).
  
3. Aplicar el método de regresión lineal simple a un indicador específico.
  
4. Aplicar métodos de estimación puntual, por intervalos y prueba de hipótesis a un indicador de salud específico.

5. Detección de datos atípicos o mal representados.
6. Creación de la matriz de correlaciones entre dos o más variables para un indicador de salud determinado.
7. Generación del mapa de correlaciones para un indicador de salud determinado.
8. Presentación del círculo de correlaciones para un indicador de salud determinado.
9. Generar a partir de los datos de un indicador de salud determinado la distribución de los datos de acuerdo a grupos con similitudes en sus datos (clústeres) y que se presenten en el plano de coordenadas de dos dimensiones.
10. Generar a partir de los datos de un indicador el dendograma respectivo que muestre las asociaciones respectivas entre sí de los datos que conforman dicho indicador.
11. Obtener la tabla de Centros de Gravedad.
12. Generar los gráficos horizontales o verticales para los centros de gravedad de un indicador de salud determinado.
13. Representar a través del diagrama de caja y bigote el conjunto de datos de un indicador de salud determinado.
14. Debe registrarse dentro de una bitácora del sistema toda acción que se realiza en él.

## 4.1.2 DESCRIPCIÓN DE REQUERIMIENTOS FUNCIONALES SOBRE FUNCIONES ESTADÍSTICAS

### 1. Resumen de medidas de posición, de tendencia central y de dispersión para un indicador de salud determinado.

Permitirá determinar deciles, cuartiles y percentiles para un conjunto de datos de una variable, también los valores de la media aritmética y la moda, así como la desviación estándar y desviación típica.

### 2. Aplicar el método de regresión lineal simple a un indicador específico.

Permitirá predecir el comportamiento de un valor (variable independiente) en función de otro valor (variable dependiente), con la finalidad de proyectar datos de los indicadores de salud.

### 3. Aplicar métodos de estimación puntual, por intervalos y prueba de hipótesis a un indicador de salud específico.

A partir de un conjunto de datos llamado muestra, se pretende, dado el conocimiento de la distribución de los datos (particularmente la Distribución Normal) y de un conjunto de estimadores y estadísticos, obtener resultados que se acerquen a lo que se puede apreciar en el universo total de una variable, mejor conocida como población.

### 4. Detección de datos atípicos o mal representados.

Determinará qué valores de los indicadores de salud, son atípicos, esto es datos que están fuera de los posibles valores de una variable, con la finalidad de obtener un mejor análisis y comprensión de los datos, ya que si existen muchos valores atípicos, se puede llegar a un análisis erróneo por parte del usuario.

### 5. Creación de la matriz de correlaciones entre dos o más variables para un indicador de salud determinado.

La correlación determina si una variable depende de otra, por tanto este requerimiento determinara el comportamiento de una variable con respecto a otras variables, su resultado se presentara en una matriz. El resultado de la matriz variará entre -1 y 1.

### 6. Generación del mapa de correlaciones para un indicador de salud determinado.

Mostrará en forma gráfica las correlaciones entre variables, lo que permitirá al usuario comprender más rápidamente el comportamiento de las variables del indicador en estudio.

**7. Presentación del círculo de correlaciones para un indicador de salud determinado.**

Determinará la correlación entre variables de forma gráfica, las variables se mostraran en forma de flechas y serán graficadas dentro de un círculo con valores entre -1 y 1. El usuario podrá determinar la correlación de variables en base a la orientación de las flechas.

**8. Generar a partir de los datos de un indicador de salud determinado la distribución de los datos de acuerdo a grupos con similitudes en sus datos (clústeres) y que se presenten en el plano de coordenadas de dos dimensiones.**

Este requerimiento permitirá al usuario separar o agrupar datos que tengan un comportamiento similar. Cada tabla de indicador contiene muchas variables, estas variables pueden ser agrupadas y clasificadas lo que permitirá encontrar patrones en los datos de los indicadores.

**9. Generar a partir de los datos de un indicador el dendograma respectivo que muestre las asociaciones respectivas entre sí de los datos que conforman dicho indicador.**

El dendograma es un tipo de representación que permite apreciar claramente las relaciones de agrupación entre los datos e incluso entre grupos de ellos aunque no las relaciones de similaridad o cercanía entre categorías. Con esta representación el usuario podrá establecer criterios de agrupación de los datos.

**10. Obtener la tabla de Centros de Gravedad.**

Se encargara de encontrar el valor promedio o más representativo de todos los datos que están agrupados dentro de un clúster de datos. Se mostrara al usuario en forma de tabla que contendrá todos los puntos medios de los clústeres encontrados.

**11. Generar los gráficos horizontales o verticales para los centros de gravedad de un indicador de salud determinado.**

Permitirá representar en forma gráfica los valores promedios de todos los datos que están agrupados en un clúster, el grafico puede ser vertical o horizontal.

**12. Representar a través del diagrama de caja y bigote el conjunto de datos de un indicador de salud determinado.**

Por medio del diagrama de caja y bigote se determinaran valores atípicos, valores máximos y mínimo, así como la mediana, cuartiles y la simetría de distribución de un conjunto de datos.

## 4.1.3 REQUERIMIENTOS NO FUNCIONALES

### Requerimientos de Calidad

Atributo de Calidad	Descripción	Actividad
<i>Seguridad</i>	Es la medida sobre la habilidad del sistema para resistir usos no autorizados mientras sigue proveyendo sus servicios a los usuarios legítimos. El desarrollo del presente módulo se adaptara a los mecanismos de seguridad que provee el Sistema Integrado de Indicadores Gerenciales.	Administración de acceso.
		Creación de respaldos y restauración.
		Inicio y cierre de sesión.
		Acceso físico y virtual al servidor.
<i>Facilidad de uso</i>	Es la facilidad de aprender a utilizar el sistema por parte de los usuarios e interactuar con sus diferentes funcionalidades.	Capacitación.
		Creación de plan de implementación.
		Creación de manuales.
<i>Facilidad de mantenimiento</i>	Es la facilidad de agregar nuevas funcionalidades al sistema, corregir errores encontrados o modificar funcionalidades y existentes.	Manual Técnico del Sistema.
		Documentación Interna del software.

Tabla 4.1. Requerimientos de calidad.

### Requerimientos de Software

A continuación en la tabla se presentan las especificaciones técnicas del software a utilizar para el desarrollo del módulo estadístico.

#### Sistema Operativo: Debian<sup>20</sup>

Versión: 8.1

Se recomienda como mínimo un Pentium 4, a 1 GHz para un sistema de escritorio.

La siguiente tabla resume los requisitos de Memoria RAM y Almacenamiento de acuerdo a dos tipos de instalación:

1. Sin escritorio, es decir, sin interfaz gráfica de usuario
2. Con escritorio, es decir, con la interfaz gráfica de usuario.

Tipo de instalación	RAM (mínimo)	RAM (recomendado)	Disco duro
Sin escritorio	64 Megabytes	256 Megabytes	1 Gigabyte
Con escritorio	128 Megabytes	512 Megabytes	5 Gigabytes

Tabla 4.2. Requisitos mínimos de sistema recomendados para Debian 8.1

<sup>20</sup> 3.4. Cumplir los requisitos mínimos de hardware. Guía de instalación de Debian GNU/Linux Visitado 13/07/2015.  
<<https://www.debian.org/releases/stable/i386/ch03s04.html.es>>

En este proyecto se opta por hacer uso de la instalación con interfaz gráfica de usuario.

### Framework de Desarrollo: Symfony<sup>21</sup>

Versión: 2.7

Como mínimo, se necesita de un servidor web, un motor de base de datos (MySQL, PostgreSQL, SQLite, o cualquier motor de base de datos compatible con PDO), y PHP 5.2.4 o posterior.

También se debe tener habilitado JSON y ctype; así como el archivo de configuración php.ini debe tener configurada correctamente la opción date.timezone

### Sistema Gestor de Base de Datos: PostgreSQL<sup>22</sup>

Versión: 9.4

Los requerimientos mínimos para instalar PostgreSQL son:

8 Megabytes de RAM

30 Megabytes de espacio en el disco duro para el código fuente

5 Megabytes de espacio en disco duro para la instalación de los ejecutables

1 Megabyte extra para las bases de datos básicas

### Lenguaje y entorno de programación para análisis estadístico y gráfico: R<sup>23</sup>

Versión: 3.2.1

Para la mayoría de los casos una computadora con Dual Core, a 2.4 GHz para un sistema de escritorio, haciendo uso principalmente de únicamente de uno de los núcleos y 2 GB de Memoria RAM. Así como 500 Megabytes de espacio en disco duro tanto para el código fuente, instalación de librerías y archivos personalizados.

### *Requerimientos de Hardware*

Los requerimientos de hardware especifican las características técnicas necesarias del equipo cliente y servidor que permitirán la implementación del producto a entregar.

---

<sup>21</sup> Documentos de Referencia de Symfony. Requerimientos para su ejecución. Texto traducido. Visitado el 13/07/2015. <<http://symfony.com/doc/current/reference/requirements.html>>

<sup>22</sup> Presentación de PostgreSQL. Requerimientos. Visitado el 13/07/2015. <<http://es.slideshare.net/etumax/postgresql-9649848>>

<sup>23</sup> Recomendaciones de Sistema para R. Texto traducido. Visitado 13/07/2015. <<https://support.rstudio.com/hc/en-us/articles/201853926-Rstudio-IDE-System-Recommendations>>

## Especificaciones Técnicas de Servidor<sup>24</sup>

En la siguiente tabla se presentan las características y especificaciones técnicas del Hardware del Servidor donde estará alojado el servidor web y el sistema gestor de base de datos que contendrán el código fuente y base de datos del sistema:

<b>Característica</b>	<b>Especificación Técnica</b>
<i>Modelo</i>	Dell PowerEdge R815
<i>Factor de Forma</i>	2U rack
<i>Procesador</i>	AMD Opteron 6100 (48 Núcleos)
<i>Socket de Procesadores</i>	4
<i>Front Side Bus / HyperTransport</i>	HyperTransport-3 Links
<i>Cache</i>	L2: 512K / Core L3: 12MB
<i>Chipset</i>	AMD
<i>Memoria RAM</i>	64 GB (32 DIMM Slots) 1333 MHz
<i>I/O Slots</i>	6 PCI-E G2 slots + storage slot: 5 8x slots 1 4x slot 1 4x storage slot
<i>Disco Duro</i>	3 TB
<i>Conexiones de red</i>	Broadcom BMC57710 10Base-T Copper Single Port NIC, PCI-E 8x
<i>Disponibilidad</i>	Hot-plug hard drives, hot-plug redundant power, dual SD modules, ECC memory, interactive LCD screen
<i>Fuente de alimentación</i>	Una fuente de alimentación de 1100W no redundante conectable en caliente Dos fuentes de alimentación de 1100W redundantes conectables en caliente
<i>Video</i>	Matrox G200eW w/ 8MB memory
<i>Soporte de Rack</i>	ReadyRail sliding rails 4-post racks
<i>Sistema Operativo</i>	Debian 7.0 Wheezy
<i>Conexiones de red</i>	Broadcom BMC57710 10Base-T Copper Single Port NIC, PCI-E 8x
<i>Disponibilidad</i>	Hot-plug hard drives, hot-plug redundant power, dual SD modules, ECC memory, interactive LCD screen
<i>Fuente de alimentación</i>	Una fuente de alimentación de 1100W no redundante conectable en caliente Dos fuentes de alimentación de 1100W redundantes conectables en caliente
<i>Video</i>	Matrox G200eW w/ 8MB memory
<i>Soporte de Rack</i>	ReadyRail sliding rails 4-post racks
<i>Sistema Operativo</i>	Debian 7.0 Wheezy

Tabla 4.3. Especificaciones técnicas de servidor.

<sup>24</sup> Datos Obtenidos de la página web del fabricante: DELL de acuerdo al modelo. <  
<http://www.dell.com/es/empresas/p/poweredge-r815/pd>>

A continuación se presentan los requerimientos mínimos necesarios de la máquina cliente, con los cuáles el sistema podrá funcionar adecuadamente:

<b>Característica</b>	<b>Especificación Técnica</b>
<i>Procesador</i>	Intel Pentium 4, 1.0 GHz
<i>Memoria RAM libre necesaria</i>	128MB
<i>Espacio libre necesario en Disco Duro</i>	1GB
<i>Conexión de red</i>	Tarjeta de Red Ethernet a 10 Gbit/s o Tarjeta Inalámbrica 802.11b
<i>Tipo de Equipo</i>	Cualquiera que posea navegador Web compatible

Tabla 4.4. Requerimientos mínimos de maquina cliente.

#### **4.1.4 RESTRICCIONES PARA EL DESARROLLO**

Las restricciones son un conjunto de limitaciones basadas en políticas organizacionales, lineamientos técnicos, estándares establecidos, etc., que rigen o sirven de guía en el desarrollo de un Sistema. A continuación se presentan las restricciones o consideraciones que el Equipo de Trabajo debe tener en cuenta para el desarrollo de proyecto, las cuales han sido dadas por la Dirección de Tecnologías de Información y Comunicaciones (DTIC).

##### *Políticas Reguladoras.*

Las Políticas Reguladoras son el conjunto de directrices que sirven para normalizar la realización de ciertas tareas o actividades. Las siguientes políticas están basadas en los lineamientos técnicos que la DTIC posee para el desarrollo y mantenimiento de Sistemas de Información, las cuales están orientadas, en su mayoría, a la utilización de sistemas operativos, software y herramientas de licenciamiento libre y a los estándares que ya se utilizan en su ambiente laboral de desarrollo y producción.

---

<sup>25</sup> Datos proporcionados como mínimos para un equipo de trabajo por la Dirección de Tecnologías de Información y Comunicaciones (DTIC).

### *Uso de Software Bajo Licenciamiento Libre.*

El software requerido para el desarrollo del proyecto y su posterior puesta en marcha, es necesario que se encuentre bajo licenciamiento libre y gratuito, exceptuándose el caso de sistemas de información anteriores a la aprobación del documento de Lineamientos técnicos<sup>26</sup> de la DTIC y circunstancias muy específicas que requieran de funcionalidades no existentes, en plataformas libres y gratuitas, las cuales deberán justificarse oportunamente en el documento de análisis.

### *Sistema Operativo.*

Para el desarrollo y producción de la Herramienta de funciones estadísticas se utilizará el Sistema Operativo Linux, en su distribución Debian 8.1, como alternativa se permitirá la utilización de Ubuntu únicamente para desarrollo del Sistema.

### *Framework de Programación.*

El Framework de programación que se utilizará para el desarrollo de la herramienta es Symfony2, en conjunto con el motor de plantillas de PHP llamado Twig y una librería encargada de ejecutar instrucciones de lenguaje R (PLR) en el Gestor de Base de Datos.

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

El SGBD que se utilizará, es PostgreSQL en su versión 9.4. Como herramienta para el manejo de la base de datos puede utilizarse PgAdmin III o phpPgadmin; sin embargo, queda a disposición del grupo de trabajo la utilización de la consola del sistema operativo con la herramienta embebida PSQL.

### *Control de Versiones.*

Para el control de versiones, se utilizará Git, que permitirá el registro de los cambios realizados sobre un archivo o conjunto de archivos a lo largo del tiempo, de modo que se pueda recuperar a versiones anteriores

---

<sup>26</sup> Ministerio de Salud. (2011). Lineamientos Técnicos para el Mantenimiento, Administración y Desarrollo de Tecnologías de Información y Comunicaciones. San Salvador, El Salvador

*Contenido Mínimo de Manuales Adjuntos al Módulo Desarrollado.*

Los manuales requeridos por la DTIC y que se entregarán junto con el Módulo Desarrollado son:

- Manual de Usuario: Este manual debe de reunir la información adecuada para que el usuario pueda conocer y utilizar adecuadamente el software desarrollado. Debe estar redactado en lenguaje entendible para el usuario final no para un usuario informático.
- Manual Técnico: Este documento debe contener todo el análisis y diseño realizado para el desarrollo del módulo de funciones estadísticas.

## 4.2 CASOS DE USO

### 4.2.1 CASO DE USO GENERAL

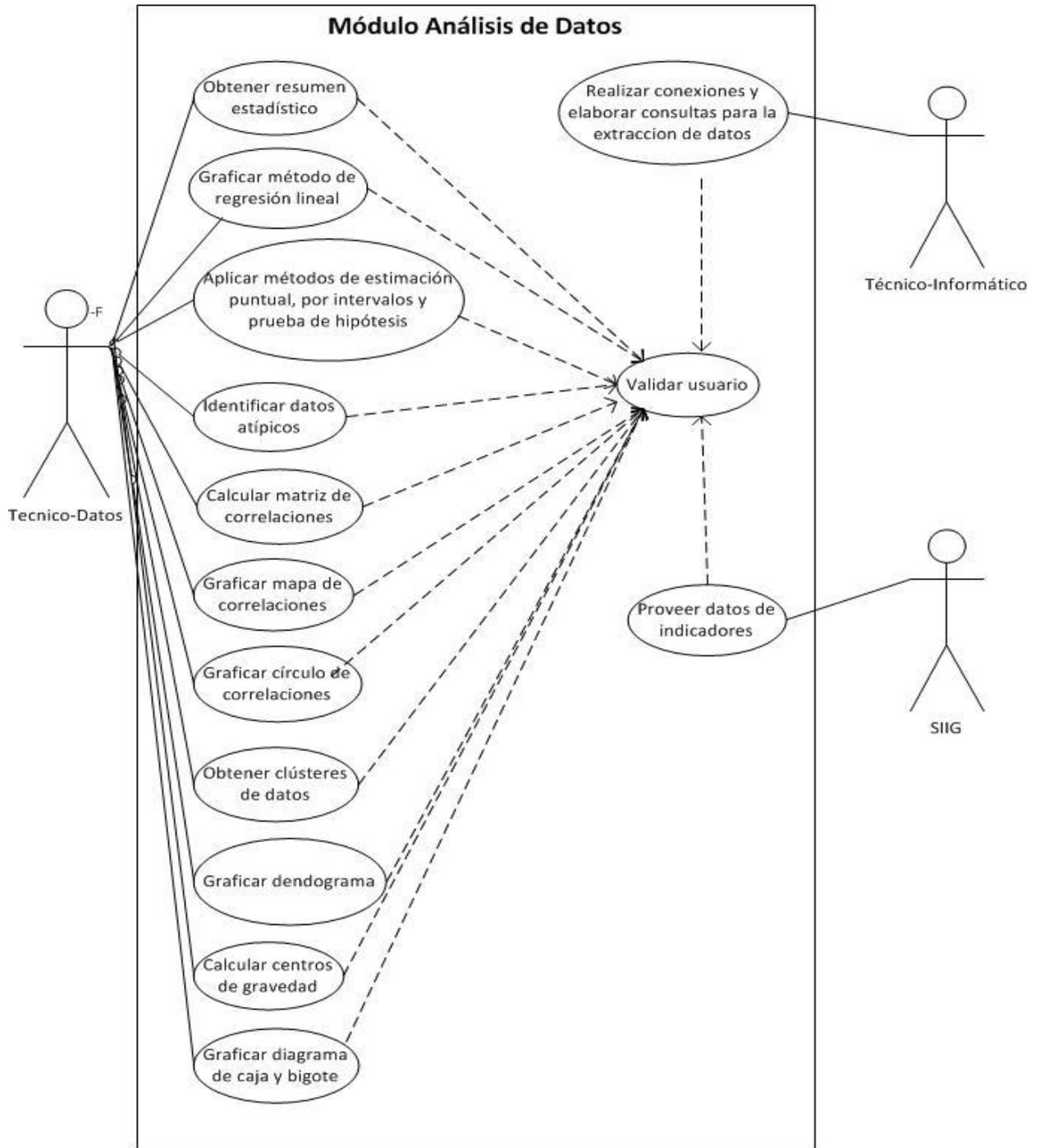


Fig.4.1-Diagrama de Caso de Uso General.

## 4.2.2 DESCRIPCIÓN CASOS DE USO

<b>Código:</b>	CU-01	<b>Nombre:</b>	Elaborar Resumen Estadístico.
<b>Actores:</b>	Técnico Informático, Técnico-Datos		
<b>Propósito:</b>	Obtener los valores estadísticos de medidas de posición, de tendencia central y de dispersión calculados a un indicador de salud.		
<b>Descripción:</b>	El usuario obtiene un resumen de valores de medidas de posición, tendencia central y dispersión de un indicador específico para su posterior análisis.		
<b>Precondiciones:</b>	El usuario ha sido habilitado para la utilización del Módulo de Análisis de Datos.		
<b>Post-Condiciones:</b>	El usuario guarda o descarta el resultado obtenido.		
<b>Flujo Básico</b>			
<b>Paso</b>	<b>Acción</b>		
<b>1</b>	Inicia caso de uso Elaborar Resumen Estadístico.		
<b>2</b>	El usuario ingresa al menú de análisis de datos.		
<b>3</b>	El usuario activa la opción de Elaborar resumen estadístico.		
<b>4</b>	El sistema carga el cuadro de dialogo de Elaborar resumen, mostrando la lista de indicadores.		
<b>5</b>	El usuario selecciona el indicador a utilizar		
<b>6</b>	El usuario establece filtros a los datos.		
<b>7</b>	El sistema muestra las variables del indicador.		
<b>8</b>	El usuario selecciona la variable a utilizar.		
<b>9</b>	El usuario selecciona el comando generar, para el cálculo de medidas de posición, tendencia central y dispersión.		
<b>10</b>	El sistema muestra en pantalla el resumen de cálculos estadísticos, para el indicador seleccionado.		
<b>11</b>	Fin del caso de uso.		
<b>Flujos Alternos</b>			
<b>Paso en el que fue invocado</b>	<b>Acción</b>		
<b>6</b>	El usuario establece filtros a los datos.		
<b>7</b>	El sistema muestra un cuadro de dialogo detallando que no existen datos para los filtros seleccionados.		
<b>8</b>	El sistema continúa con el paso 4 del flujo básico.		
<b>Frecuencia Esperada:</b>	Mensual		
<b>Importancia:</b>	Moderada.		
<b>Urgencia:</b>	Moderada.		

Tabla 4.5. Caso de Uso Resumen Estadístico.

<b>Código:</b>	CU-02	<b>Nombre:</b>	Método de Regresión Lineal Simple
<b>Actores:</b>	Técnico informático, Técnico-datos		
<b>Propósito:</b>	Obtener un modelo de regresión lineal simple.		
<b>Descripción:</b>	El usuario obtiene un modelo de regresión lineal simple, siempre que los datos cumplan con dicho modelo.		
<b>Precondiciones:</b>	El usuario ha sido habilitado para la utilización del Módulo de Análisis de Datos.		
<b>Post-Condiciones:</b>	El usuario guarda o descarta el resultado obtenido.		
<b>Flujo Básico</b>			
<b>Paso</b>	<b>Acción</b>		
<b>1</b>	Inicia caso de uso Regresión Lineal Simple.		
<b>2</b>	El usuario ingresa al menú de análisis de datos.		
<b>3</b>	El usuario activa la opción de Regresión Lineal Simple.		
<b>4</b>	El sistema carga el cuadro de dialogo de Regresión Lineal Simple, mostrando la lista de indicadores.		
<b>5</b>	El usuario selecciona el indicador a utilizar.		
<b>6</b>	El usuario selecciona un rango de fecha de los datos.		
<b>7</b>	El sistema muestra las variables del indicador.		
<b>8</b>	El usuario selecciona las variables a utilizar.		
<b>9</b>	El usuario selecciona el comando generar.		
<b>10</b>	El sistema muestra en pantalla el grafico de regresión lineal.		
<b>11</b>	Fin del caso de uso.		
<b>Flujos Alternos</b>			
<b>Paso en el que fue invocado</b>	<b>Acción</b>		
<b>6</b>	El usuario selecciona un rango de fecha de los datos.		
<b>7</b>	El sistema muestra un cuadro de dialogo detallando que no existen datos para las fechas seleccionadas.		
<b>8</b>	El sistema continúa con el paso 4 del flujo básico.		
<b>Frecuencia Esperada:</b>	Mensual		
<b>Importancia:</b>	Moderada.		
<b>Urgencia:</b>	Moderada.		

Tabla 4.6. Caso de Uso Regresión Lineal.

<b>Código:</b>	CU-03	<b>Nombre:</b>	Métodos de estimación puntual, por intervalos y prueba de hipótesis.
<b>Actores:</b>	Técnico-datos		
<b>Propósito:</b>	Aplicar métodos de inferencia estadística a muestras de datos obtenidos de indicadores de salud.		
<b>Descripción:</b>	El usuario obtiene resultados muestrales sobre datos de indicadores de salud.		
<b>Precondiciones:</b>	El usuario ha sido habilitado para la utilización del Módulo de Análisis de Datos.		
<b>Post-Condiciones:</b>	El usuario guarda o descarta el resultado obtenido.		
<b>Flujo Básico</b>			
<b>Paso</b>	<b>Acción</b>		
1	Inicia caso de uso.		
2	El usuario ingresa al menú de análisis de datos.		
3	El usuario activa la opción de Métodos de inferencia estadística.		
4	El sistema carga el cuadro de dialogo Métodos de inferencia estadística, mostrando la lista de indicadores, así como filtros de fecha para los datos.		
5	El usuario selecciona el indicador a utilizar y establece filtros de fecha.		
6	El sistema muestra las variables del indicador.		
7	El usuario selecciona las variables a utilizar.		
8	El sistema muestra un cuadro de dialogo donde se muestran los métodos de inferencia estadística disponibles.		
9	El usuario selecciona cualquiera de los métodos estadísticos de inferencia y digita nivel de confianza o valor de hipótesis alternativa o ambas dependiendo del método a utilizar.		
10			
11	El usuario selecciona el comando calcular, para calcular el método de inferencia seleccionado.		
12	El sistema muestra en pantalla el resultado del cálculo.		
13	Fin del caso de uso.		
<b>Flujos Alternos</b>			
<b>Paso en el que fue invocado</b>	<b>Acción</b>		
5	El usuario selecciona el indicador a utilizar y establece filtros de fecha.		
6	El sistema muestra un cuadro de dialogo detallando que no existen datos para las fechas ingresadas.		
7	El sistema continúa con el paso 4 del flujo básico.		
<b>Frecuencia Esperada:</b>	Mensual		
<b>Importancia:</b>	Moderada.		
<b>Urgencia:</b>	Moderada.		

Tabla 4.7. Caso de Uso Métodos Inferenciales.

<b>Código:</b>	CU-04	<b>Nombre:</b>	Detectar Datos Atípicos
<b>Actores:</b>	Técnico- informático, Técnico-datos		
<b>Propósito:</b>	Identificar los datos mal representados de un indicador específico.		
<b>Descripción:</b>	El usuario obtiene los datos atípicos de un indicador específico para su posterior análisis.		
<b>Precondiciones:</b>	El usuario ha sido habilitado para la utilización del Módulo de Análisis de Datos.		
<b>Post-Condiciones:</b>	El usuario guarda o descarta el resultado obtenido.		
<b>Flujo Básico</b>			
<b>Paso</b>	<b>Acción</b>		
1	Inicia caso de uso Detectar Datos Atípicos.		
2	El usuario ingresa al menú de análisis de datos.		
3	El usuario activa la opción de Datos Atípicos.		
4	El sistema carga el cuadro de dialogo de Datos Atípicos, mostrando la lista de indicadores.		
5	El usuario selecciona el indicador a utilizar.		
6	El sistema muestra las variables del indicador.		
7	El usuario selecciona las variables a utilizar, así como también establece filtros de fecha a los datos.		
8	El usuario selecciona el comando generar, para la identificación de datos atípicos.		
9	El sistema muestra en pantalla los datos atípicos encontrados del indicador de salud seleccionado.		
10	Fin del caso de uso.		
<b>Flujos Alternos</b>			
<b>Paso en el que fue invocado</b>	<b>Acción</b>		
7	El usuario selecciona las variables a utilizar, así como también establece filtros de fecha a los datos.		
8	El sistema muestra un cuadro de dialogo detallando que no existen datos para las variables y filtros seleccionados.		
9	El sistema continúa con el paso 4 del flujo básico.		
<b>Frecuencia Esperada:</b>	Mensual		
<b>Importancia:</b>	Moderada.		
<b>Urgencia:</b>	Moderada.		

Tabla 4.8. Caso de Uso Detección Datos Atípicos.

<b>Código:</b>	CU-05	<b>Nombre:</b>	Calcular Matriz de Correlaciones
<b>Actores:</b>	Técnico informático, Técnico-datos		
<b>Propósito:</b>	Obtener la matriz de correlaciones.		
<b>Descripción:</b>	El usuario obtiene la matriz de correlaciones de un indicador específico para su posterior análisis.		
<b>Precondiciones:</b>	El usuario ha sido habilitado para la utilización del Módulo de Análisis de Datos.		
<b>Post-Condiciones:</b>	El usuario guarda o descarta el resultado obtenido.		
<b>Flujo Básico</b>			
<b>Paso</b>	<b>Acción</b>		
1	Inicia caso de uso Calcular Matriz de Correlaciones.		
2	El usuario ingresa al menú de análisis de datos.		
3	El usuario activa la opción de Matriz de Correlaciones.		
4	El sistema carga la pantalla de Matriz de Correlaciones, mostrando la lista de indicadores, así como filtros para los datos.		
5	El usuario selecciona el indicador a utilizar y establece filtros de fecha.		
6	El sistema muestra las variables del indicador.		
7	El usuario selecciona las variables a utilizar.		
8	El usuario selecciona el comando generar, para la generación de la matriz de correlaciones.		
9	El sistema muestra en pantalla la matriz de correlaciones calculada al indicador de salud seleccionado.		
10	Fin del caso de uso.		
<b>Flujos Alternos</b>			
<b>Paso en el que fue invocado</b>	<b>Acción</b>		
5	El usuario selecciona el indicador a utilizar y establece filtros de fecha.		
6	El sistema muestra un cuadro de dialogo detallando que no existen datos para las fechas seleccionadas.		
7	El sistema continúa con el paso 4 del flujo básico.		
<b>Frecuencia Esperada:</b>	Mensual		
<b>Importancia:</b>	Moderada.		
<b>Urgencia:</b>	Moderada.		

Tabla 4.9. Caso de Uso Calcular Matriz de Correlaciones.

<b>Código:</b>	CU-06	<b>Nombre:</b>	Graficar Mapa de Correlaciones
<b>Actores:</b>	Técnico-informático, Técnico-datos		
<b>Propósito:</b>	Obtener Mapa de Correlaciones		
<b>Descripción:</b>	El usuario obtiene el mapa de correlaciones a un indicador específico para su posterior análisis.		
<b>Precondiciones:</b>	El usuario ha sido habilitado para la utilización del Módulo de Análisis de Datos.		
<b>Post-Condiciones:</b>	El usuario guarda o descarta el resultado obtenido.		
<b>Flujo Básico</b>			
<b>Paso</b>	<b>Acción</b>		
<b>1</b>	Inicia caso de uso Graficar Mapa de Correlaciones.		
<b>2</b>	El usuario ingresa al menú de análisis de datos.		
<b>3</b>	El usuario activa la opción de Mapa de Correlaciones.		
<b>4</b>	El sistema carga la pantalla de Mapa de Correlaciones, mostrando la lista de indicadores, así como filtros para los datos.		
<b>5</b>	El usuario selecciona el indicador a utilizar y establece filtros de fecha.		
<b>6</b>	El sistema muestra las variables del indicador.		
<b>7</b>	El usuario selecciona las variables a utilizar.		
<b>8</b>	El usuario selecciona el comando generar, para la generación del mapa de correlaciones.		
<b>9</b>	El sistema muestra en pantalla el mapa de correlaciones calculada al indicador de salud seleccionado		
<b>10</b>	Fin del caso de uso.		
<b>Flujos Alternos</b>			
<b>Paso en el que fue invocado</b>	<b>Acción</b>		
<b>5</b>	El usuario selecciona el indicador a utilizar y establece filtros de fecha.		
<b>6</b>	El sistema muestra un cuadro de dialogo detallando que no existen datos para las fechas seleccionadas.		
<b>7</b>	El sistema continúa con el paso 4 del flujo básico.		
<b>Frecuencia Esperada:</b>	Mensual		
<b>Importancia:</b>	Moderada.		
<b>Urgencia:</b>	Moderada.		

Tabla 4.10. Graficar Mapa de Correlaciones.

<b>Código:</b>	CU-07	<b>Nombre:</b>	Graficar Circulo de Correlaciones
<b>Actores:</b>	Técnico-informático, Técnico-datos		
<b>Propósito:</b>	Obtener el Círculo de Correlaciones.		
<b>Descripción:</b>	El usuario obtiene el círculo de correlaciones a un indicador específico para su posterior análisis.		
<b>Precondiciones:</b>	El usuario ha sido habilitado para la utilización del Módulo de Análisis de Datos.		
<b>Post-Condiciones:</b>	El usuario guarda o descarta el resultado obtenido.		
<b>Flujo Básico</b>			
<b>Paso</b>	<b>Acción</b>		
<b>1</b>	Inicia caso de uso Graficar Círculo de Correlaciones.		
<b>2</b>	El usuario ingresa al menú de análisis de datos.		
<b>3</b>	El usuario activa la opción de Círculo de Correlaciones.		
<b>4</b>	El sistema carga la pantalla de Círculo de Correlaciones, mostrando la lista de indicadores, así como filtros para los datos.		
<b>5</b>	El usuario selecciona el indicador a utilizar.		
<b>6</b>	El sistema muestra las variables del indicador.		
<b>7</b>	El usuario selecciona las variables a utilizar, así como también establece filtros a los datos.		
<b>8</b>	El usuario selecciona el comando generar, para la generación del círculo de correlaciones.		
<b>9</b>	El sistema muestra en pantalla el círculo de correlaciones calculada al indicador de salud seleccionado.		
<b>10</b>	Fin del caso de uso		
<b>Flujos Alternos</b>			
<b>Paso en el que fue invocado</b>	<b>Acción</b>		
<b>7</b>	El usuario selecciona las variables a utilizar, así como también establece filtros a los datos.		
<b>8</b>	El sistema muestra un cuadro de dialogo detallando que no existen datos para las variables y filtros seleccionados.		
<b>9</b>	El sistema continúa con el paso 4 del flujo básico.		
<b>Frecuencia Esperada:</b>	Mensual		
<b>Importancia:</b>	Moderada.		
<b>Urgencia:</b>	Moderada.		

Tabla 4.11. Caso de Uso Graficar Círculo de Correlaciones.

<b>Código:</b>	CU-08	<b>Nombre:</b>	Graficar clústeres de datos
<b>Actores:</b>	Técnico-informático, Técnico-datos		
<b>Propósito:</b>	Obtener clústeres de datos y graficarlos en dos dimensiones.		
<b>Descripción:</b>	El usuario obtiene un gráfico en dos dimensiones de los clústeres encontrados a un indicador de salud.		
<b>Precondiciones:</b>	El usuario ha sido habilitado para la utilización del Módulo de Análisis de Datos.		
<b>Post-Condiciones:</b>	El usuario guarda o descarta el resultado obtenido.		
<b>Flujo Básico</b>			
<b>Paso</b>	<b>Acción</b>		
<b>1</b>	Inicia caso de uso Graficar clústeres de datos.		
<b>2</b>	El usuario ingresa al menú de análisis de datos.		
<b>3</b>	El usuario activa la opción de Clústeres de datos.		
<b>4</b>	El sistema carga un cuadro de dialogo, mostrando la lista de indicadores, así como filtros para los datos.		
<b>5</b>	El usuario selecciona el indicador a utilizar.		
<b>6</b>	El sistema muestra las variables del indicador.		
<b>7</b>	El usuario selecciona las variables a utilizar, así como también establece filtros de fecha a los datos.		
<b>8</b>	El usuario selecciona el comando generar, para la generación del gráfico.		
<b>9</b>	El sistema muestra en pantalla el gráfico generado del indicador de salud seleccionado.		
<b>10</b>	Fin del caso de uso.		
<b>Flujos Alternos</b>			
<b>Paso en el que fue invocado</b>	<b>Acción</b>		
<b>7</b>	El usuario selecciona las variables a utilizar, así como también establece filtros de fecha a los datos.		
<b>8</b>	El sistema muestra un cuadro de dialogo detallando que no existen datos para las fechas seleccionadas.		
<b>9</b>	El sistema continúa con el paso 4 del flujo básico.		
<b>Frecuencia Esperada:</b>	Mensual		
<b>Importancia:</b>	Moderada.		
<b>Urgencia:</b>	Moderada.		

Tabla 4.12. Caso de Uso Graficar Clústeres de Datos.

<b>Código:</b>	CU-09	<b>Nombre:</b>	Graficar Dendograma
<b>Actores:</b>	Técnico-informático, Técnico-datos		
<b>Propósito:</b>	Visualizar el Dendograma de un indicador.		
<b>Descripción:</b>	El usuario obtiene el Dendograma de un indicador específico para su posterior análisis.		
<b>Precondiciones:</b>	El usuario ha sido habilitado para la utilización del Módulo de Análisis de Datos.		
<b>Post-Condicion:</b>	El usuario guarda o descarta el resultado obtenido.		
<b>Flujo Básico</b>			
<b>Paso</b>	<b>Acción</b>		
<b>1</b>	Inicia caso de uso Graficar Dendograma.		
<b>2</b>	El usuario ingresa al menú de análisis de datos.		
<b>3</b>	El usuario activa la opción de Dendograma.		
<b>4</b>	El sistema carga la pantalla de Dendograma, mostrando la lista de indicadores, así como filtros para los datos.		
<b>5</b>	El usuario selecciona el indicador a utilizar		
<b>6</b>	El sistema muestra las variables del indicador.		
<b>7</b>	El usuario selecciona las variables a utilizar, así como también establece filtros de fecha los datos.		
<b>8</b>	El usuario selecciona el comando generar, para la generación del Dendograma.		
<b>9</b>	El sistema muestra en pantalla el Dendograma generado del indicador de salud seleccionado.		
<b>10</b>	Fin del caso de uso		
<b>Flujos Alternos</b>			
<b>Paso en el que fue invocado</b>	<b>Acción</b>		
<b>7</b>	El usuario selecciona las variables a utilizar, así como también establece filtros de fecha los datos.		
<b>8</b>	El sistema muestra un cuadro de dialogo detallando que no existen datos para las fechas seleccionadas.		
<b>9</b>	El sistema continúa con el paso 4 del flujo básico.		
<b>Frecuencia Esperada:</b>	Mensual		
<b>Importancia:</b>	Moderada.		
<b>Urgencia:</b>	Moderada.		

Tabla 4.13. Caso de Uso Graficar Dendograma.

<b>Código:</b>	CU-10	<b>Nombre:</b>	Calcular Centros de Gravedad
<b>Actores:</b>	Técnico-informático, Técnico-datos		
<b>Propósito:</b>	Obtener los centros de gravedad de un indicador específico.		
<b>Descripción:</b>	El usuario obtiene el cálculo de los centros de gravedad de un indicador específico para su posterior análisis.		
<b>Precondiciones:</b>	El usuario ha sido habilitado para la utilización del Módulo de Análisis de Datos.		
<b>Post-Condiciones:</b>	El usuario guarda o descarta el resultado obtenido.		
<b>Flujo Básico</b>			
<b>Paso</b>	<b>Acción</b>		
1	Inicia caso de uso Calcular Centros de Gravedad.		
2	El usuario ingresa al menú de análisis de datos.		
3	El usuario activa la opción de Centros de Gravedad.		
4	El sistema carga la pantalla de Centros de Gravedad, mostrando la lista de indicadores, así como filtros para los datos.		
5	El usuario selecciona el indicador a utilizar		
6	El sistema muestra las variables del indicador.		
7	El usuario selecciona las variables a utilizar, así como también establece filtros a los datos.		
8	El usuario selecciona el tipo de grafico que puede ser vertical u horizontal.		
9	El usuario selecciona el comando generar, para la generación de los Centros de Gravedad.		
10	El sistema muestra en pantalla los Centros de gravedad calculados del indicador de salud seleccionado.		
11	Fin del caso de uso.		
<b>Flujos Alternos</b>			
<b>Paso en el que fue invocado</b>	<b>Acción</b>		
7	El usuario selecciona las variables a utilizar, así como también establece filtros a los datos.		
8	El sistema muestra un cuadro de dialogo detallando que no existen datos para las variables y filtros seleccionados.		
9	El sistema continúa con el paso 4 del flujo básico.		
<b>Frecuencia Esperada:</b>	Mensual		
<b>Importancia:</b>	Moderada.		
<b>Urgencia:</b>	Moderada.		

Tabla 4.14. Caso de Uso Calcular Centros de Gravedad.

<b>Código:</b>	CU-11	<b>Nombre:</b>	Graficar Diagrama de Caja y Bigote
<b>Actores:</b>	Técnico-informático, Técnico-datos		
<b>Propósito:</b>	Obtener el diagrama de caja y bigote de un indicador específico.		
<b>Descripción:</b>	El usuario obtiene el diagrama de caja y bigote de un indicador específico para su posterior análisis.		
<b>Precondiciones:</b>	El usuario ha sido habilitado para la utilización del Módulo de Análisis de Datos.		
<b>Post-Condiciones:</b>	El usuario guarda o descarta el resultado obtenido.		
<b>Flujo Básico</b>			
<b>Paso</b>	<b>Acción</b>		
<b>1</b>	Inicia caso de uso Generar Diagrama de Caja y Bigote		
<b>2</b>	El usuario ingresa al menú de análisis de datos.		
<b>3</b>	El usuario activa la opción de Diagrama de caja y bigote.		
<b>4</b>	El sistema carga el cuadro de dialogo de Diagrama de caja y bigote, mostrando la lista de indicadores, así como filtros para los datos.		
<b>5</b>	El usuario selecciona el indicador a utilizar.		
<b>6</b>	El sistema muestra las variables del indicador.		
<b>7</b>	El usuario selecciona las variables a utilizar, así como también establece filtros a los datos.		
<b>8</b>	El usuario selecciona el comando generar, para la generación del Diagrama de caja y bigote		
<b>9</b>	El sistema muestra en pantalla el Diagrama de Caja y Bigote generado del indicador de salud seleccionado.		
<b>10</b>	Fin del caso de uso.		
<b>Flujos Alternos</b>			
<b>Paso en el que fue invocado</b>	<b>Acción</b>		
<b>7</b>	El usuario selecciona las variables a utilizar, así como también establece filtros a los datos.		
<b>8</b>	El sistema muestra un cuadro de dialogo detallando que no existen datos para las variables y filtros seleccionados.		
<b>9</b>	El sistema continúa con el paso 4 del flujo básico.		
<b>Frecuencia Esperada:</b>	Mensual		
<b>Importancia:</b>	Moderada.		
<b>Urgencia:</b>	Moderada.		

Tabla 4.15. Caso de Uso Graficar Diagrama de Caja y Bigote.

### **4.3 ELABORACIÓN DIAGRAMA ENTIDAD RELACIÓN**

En el desarrollo del presente proyecto se tomara como base el diagrama entidad relación del Sistema de Indicadores Gerenciales, por las siguientes razones:

- Es una herramienta que se integrara al Sistema de Indicadores Gerenciales (SIIG).
- El análisis de datos se aplicara a los almacenes de datos que genera el SIIG.
- Los almacenes de datos son tablas donde se reúne información de todas las bases de datos del MINSAL y se utilizan para poder calcular los indicadores de salud, que son la base para poder aplicar las funciones estadísticas.

El módulo permitirá guardar los resultados de los análisis, por lo que se agregaran algunas entidades nuevas al diagrama actual del sistema.

### 4.3.1 DIAGRAMA ENTIDAD RELACION DEL SISTEMA DE INDICADORES

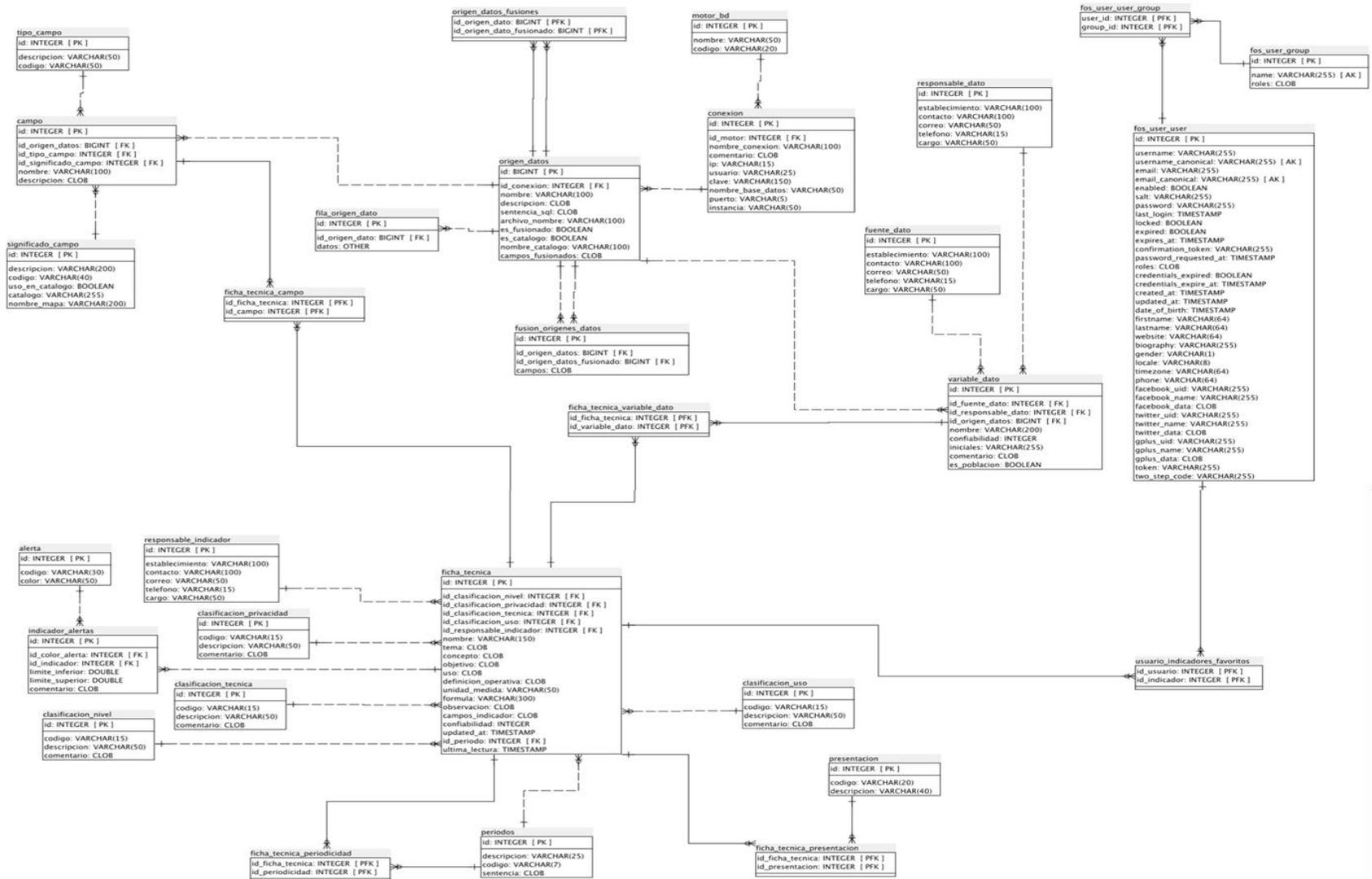


Fig. 4.2- Diagrama entidad relación del SIIG

## 4.3.2 DIAGRAMA ENTIDAD RELACIÓN MÓDULO DE ANALISIS DE DATOS

A continuación se presenta el diagrama Entidad Relación para el Modulo de Análisis de Datos:

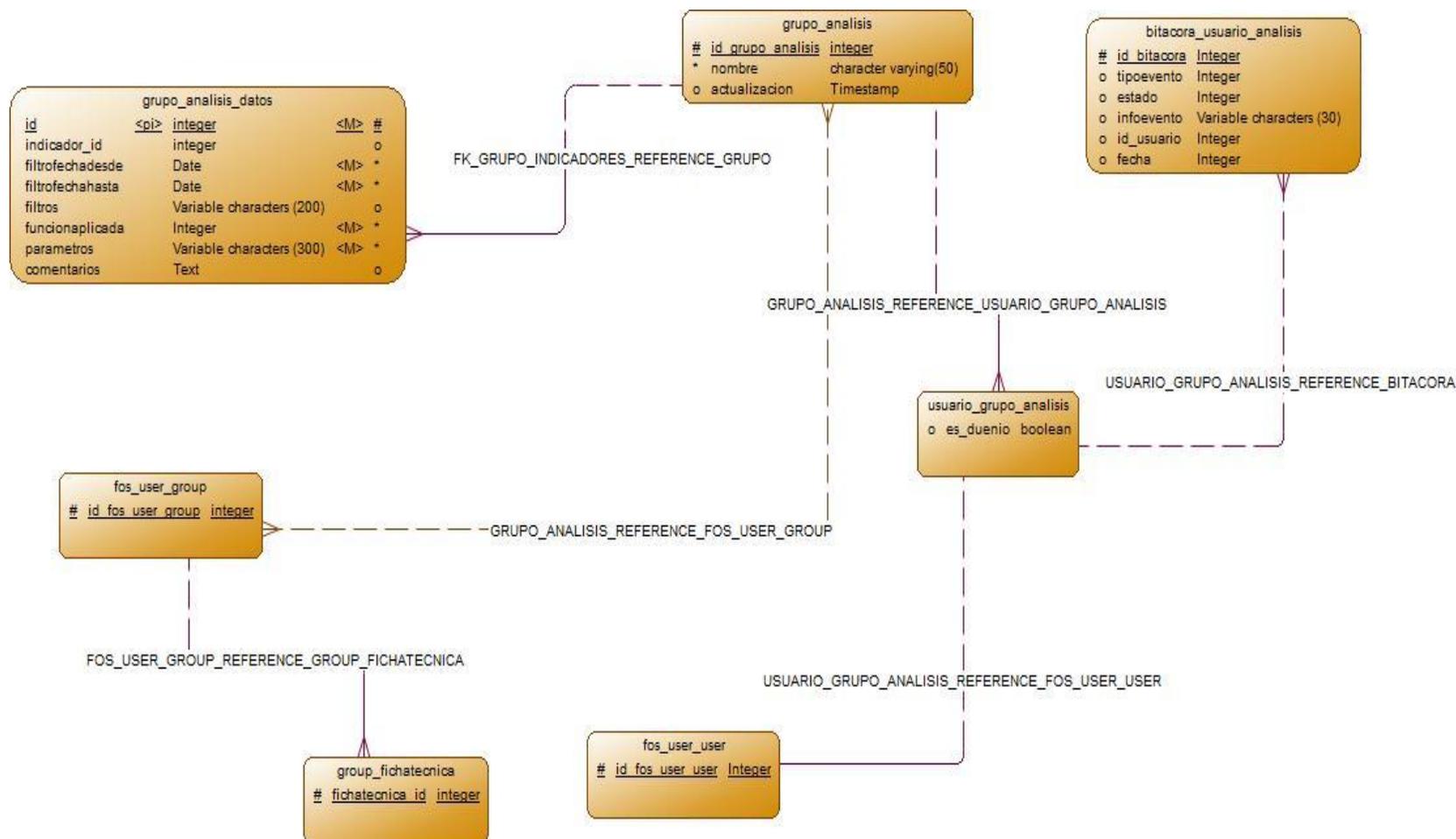


Fig. 4.3- Diagrama entidad relación módulo de funciones estadísticas.

**Observaciones:** Las entidades *fos\_user\_group*, *fos\_user\_user* son parte del sistema actual (SIIG/eTAB) por lo que no están bien detalladas, se han colocado en este modelo (Fig. 4.3) para comprender que se deben hacer las relaciones respectivas con dichas entidades.

## **4.4 PROCEDIMIENTO PARA INTEGRAR FUNCIONES ESTADÍSTICAS.**

### **1. Propósito**

Describir los pasos a seguir para desarrollar e integrar funciones estadísticas sobre las variables de los indicadores de salud para el análisis de datos en el Sistema Integrado de Indicadores (SIIG/eTAB).

### **2. Alcance**

Este procedimiento es aplicable a todas las variables almacenadas dentro del Almacén de Datos del Sistema Integrado de Indicadores Gerenciales (SIIG/eTAB).

### **3. Políticas de Operación**

- i. Solo se aplicara sobre las variables almacenadas dentro del almacén de datos del SIIG/eTAB.
- ii. Es responsabilidad de la Dirección de Tecnologías de Información y Comunicaciones (DTIC) el adecuado mantenimiento y actualización de las variables almacenadas dentro del almacén de datos del SIIG/eTAB.
- iii. Es responsabilidad del comité de indicadores la generación de nuevos indicadores que deban ser incorporados dentro del SIIG/eTAB.
- iv. Es responsabilidad del equipo de desarrollo informático el cumplir con los resultados esperados de las funciones estadísticas que se van a aplicar sobre las variables de los indicadores de salud.

#### 4. Diagrama de Procedimiento.

##### 4.1 Diagrama de procedimiento Elección de Indicadores de Salud.

Los usuarios asignados por la DTIC como consultores internos para el desarrollo de este proyecto elegirán un conjunto de indicadores de salud.

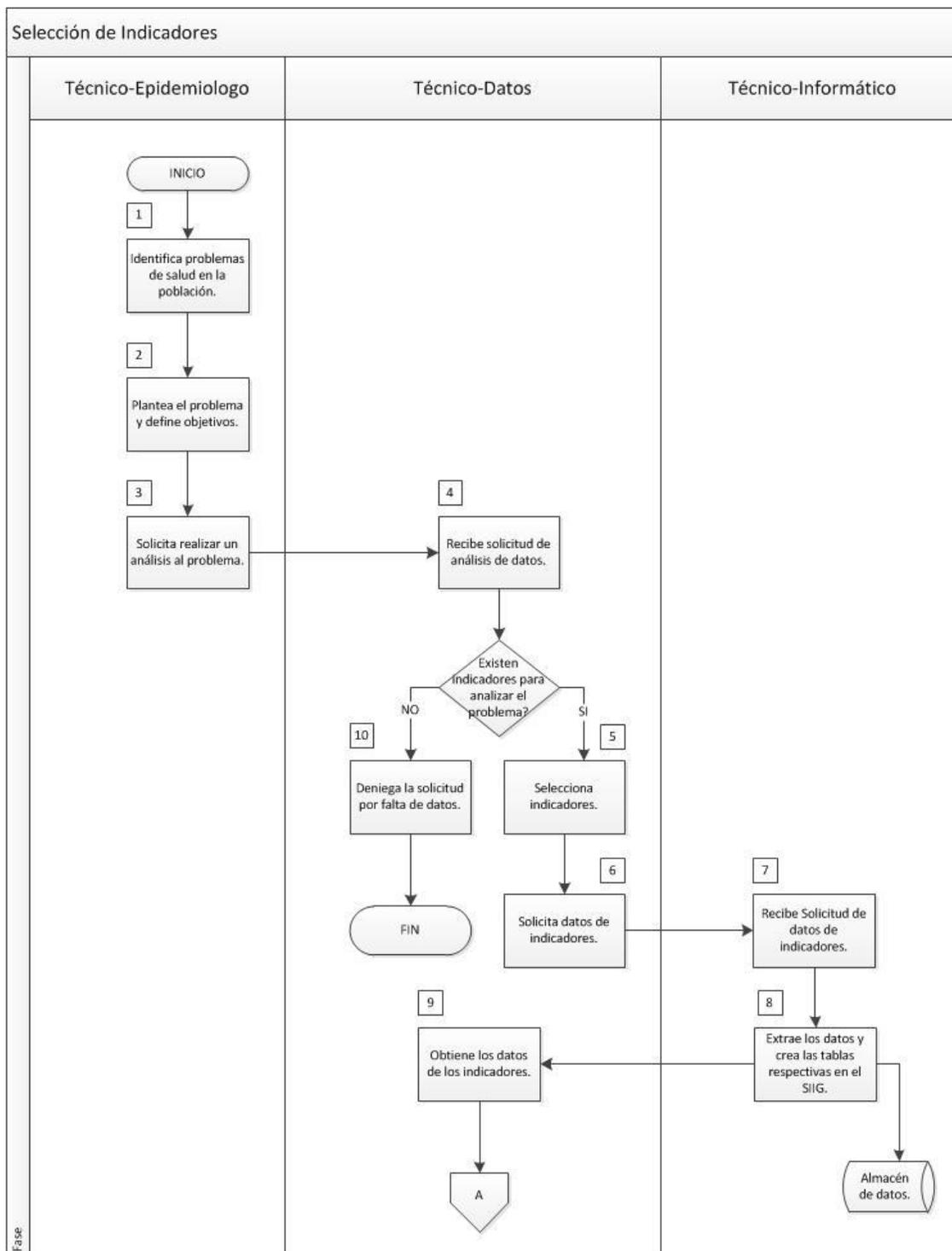


Fig.4.4- Selección de Indicadores.

#### 4.1.1 Descripción del procedimiento Elección de Indicadores de Salud

<b>Secuencia de Etapas</b>	<b>Actividad</b>	<b>Responsable</b>
1. Identifica problemas de salud en la población.	1.1 Investiga y verifica últimos resultados en cuanto a salud de la población se refiere lo que le permite identificar problemas de salud.	Técnico-Epidemiólogo
2. Plantea el problema y define los objetivos.	2.1 Plantea el problema resultado de la investigación realizada. 2.2 Define objetivos a alcanzar de acuerdo al problema en estudio.	Técnico-Epidemiólogo
3. Solicita realizar un análisis al problema.	3.1 Elabora una solicitud de análisis al problema, donde detalla el problema a analizar, los objetivos a cumplir.	Técnico-Epidemiólogo
4. Recibe solicitud de análisis de datos.	4.1 Obtiene la solicitud de análisis de datos y verifica su contenido.	Técnico-Datos
5. Selecciona indicadores o deniega la solicitud por falta de datos.	5.1 Verifica si existen los indicadores necesarios para analizar el problema y los selecciona, si no existen deniega la solicitud.	Técnico-Datos
6. Solicita datos de indicadores.	6.1 Elabora una solicitud de datos de indicadores detallando cuales datos y variables son necesarios para el estudio.	Técnico-Datos
7. Recibe solicitud de datos de indicadores	7.1 Recibe la solicitud y procede a revisar su contenido.	Técnico-Informático
8. Extrae los datos y crea las tablas respectivas en el SIIG.	8.1 Crea una sentencia SQL para la extracción de datos. 8.2 Ejecuta la sentencia SQL.	Técnico-Informático
9. Obtiene los datos de los indicadores.	9.1 Verifica en el SIIG, si ya están disponibles los datos solicitados.	Técnico-Datos

Tabla 4.16. Selección de Indicadores.

## 4.2 Diagrama de procedimiento Análisis de Datos de los Indicadores.

Se tomarán los datos contenidos dentro del Almacén de datos del Ministerio de Salud que pertenezcan a los indicadores de salud previamente seleccionados para analizarse con el objetivo de conocer las fuentes que dan origen a cada uno de los indicadores de salud y sus características principales (tipo de dato, longitud, restricciones, entre otras).

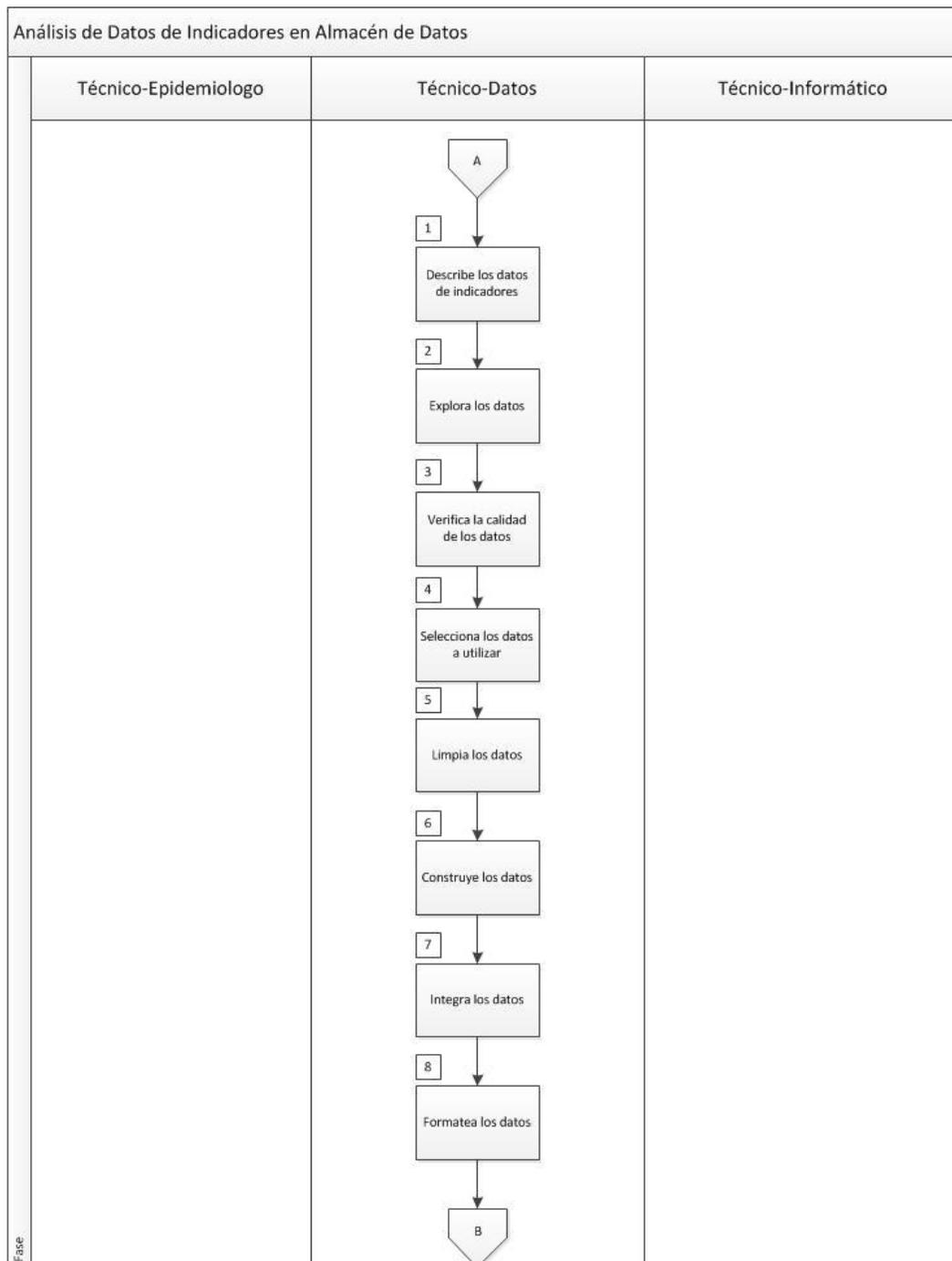


Fig.4.5- Análisis de Datos de Indicadores.

#### 4.2.1 Descripción de Procedimiento de Análisis de Datos de Indicadores en Almacén de Datos

Secuencia de Etapas	Actividad	Responsable
1. Describe los datos de indicadores.	1.1 Identificar tablas y sus relaciones. 1.2 Comprobar el volumen de datos y su complejidad. 1.3 Identificar si los datos contienen entradas de texto libres. 1.4 Comprobar los tipos de atributos y su rango de valores. 1.5 Comprender el significado de cada atributo. 1.6 Realizar un análisis estadístico básico a las variables involucradas y analizar los resultados. 1.7 Decidir si las variables son relevantes para el objetivo que se busca.	Técnico-Datos
2. Explorar los datos.	2.1 Analizar en detalle las propiedades de variables sobresalientes. 2.2 Identificar características de subpoblaciones.	Técnico-Datos
3. Verificar la calidad de los datos.	3.1 Comprobar si todos los valores posibles están representados. 3.2 Verificar que el significado de las variables y sus valores se correspondan simultáneamente. 3.3 Identificar valores atípicos. 3.4 Comprobar la ortografía y formato de valores. 3.4 Si se usan archivos planos verificar la calidad de los datos.	Técnico-Datos

<b>Secuencia de Etapas</b>	<b>Actividad</b>	<b>Responsable</b>
4. Selecciona los datos a utilizar	<p>4.1 Seleccionar variables a utilizar basándose en la calidad de datos y exploración de datos.</p> <p>4.2 Seleccionar diferentes subconjuntos de datos.</p> <p>4.3 Documentar el razonamiento para la inclusión o exclusión de datos.</p>	Técnico-Datos
5. Limpia los datos	<p>5.1 Reconsiderar como tratar cualquier valor erróneo encontrado.</p> <p>5.2 Corregir, remover o ignorar valores erróneos.</p> <p>5.3 Decidir cómo tratar valores especiales.</p> <p>5.4 Realizar una nueva selección de datos si es necesario.</p>	Técnico-Datos
6. Construye los datos.	<p>6.1 Verificar si cualquier atributo puede ser normalizado.</p> <p>6.2 Considerar agregar nuevos atributos derivados.</p> <p>6.3 Agregar atributos nuevos a los datos de acceso.</p>	Técnico-Datos
7. Integrar los datos.	7.1 Integrar las diferentes fuentes de datos para crear nuevos registros.	Técnico-Datos
8. Formatea los datos.	8.1 Reordenar los datos de acuerdo a los requerimientos de la aplicación a utilizar.	Técnico-Datos

Tabla 4.17. Análisis de Datos de Indicadores.

### 4.3 Diagrama de procedimiento Análisis de Formulas Estadísticas aplicables a los datos.

Una vez analizados los datos, se elegirán un conjunto de fórmulas estadísticas aplicables que sean las más adecuadas con los resultados que se desean obtener y se analizarán con el objetivo de conocer sus características principales (datos de entrada, datos de salida, parámetros aceptados, entre otros).

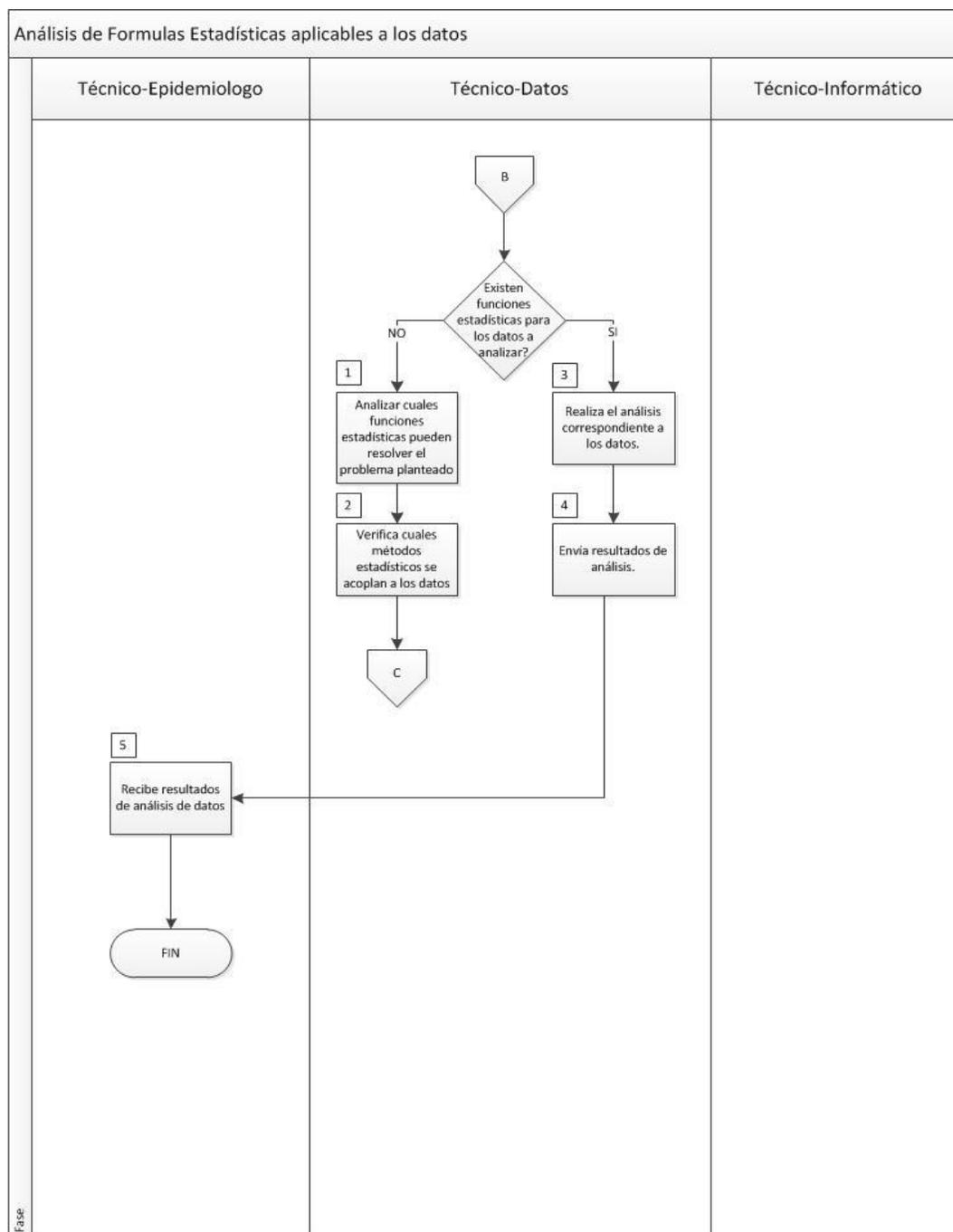


Fig.4.6- Análisis de fórmulas estadísticas.

### 4.3.1 Descripción de Procedimiento Análisis de Formulas Estadísticas Aplicables a los Datos.

Secuencia de Etapas	Actividad	Responsable
1. Analizar cuales funciones estadísticas pueden resolver el problema planteado.	1.1 Identificar y listar funciones estadísticas en R. 1.2 Identificar tipo de variables que soportan las funciones. 1.3 Identificar parámetros requeridos por las funciones.	Técnico-Datos
2. Verifica cuales métodos estadísticos se acoplan a los datos.	2.1 Probar las funciones estadísticas con subconjuntos de datos. 2.2 Verificar los resultados y analizar comportamiento de los datos.	Técnico-Datos
3. Realiza el análisis correspondiente a los datos.	3.1 Ejecuta las funciones estadísticas a los datos proporcionados. 3.2 Analiza los resultados obtenidos. 3.3 Elabora un informe con los resultados obtenidos.	Técnico-Datos
4. Envía resultados de análisis.	4.1 Envía un informe detallando los resultados obtenidos.	Técnico-Datos
5. Recibe resultados de análisis.	5.1 Obtiene los resultados del análisis de datos y elabora sus conclusiones.	Técnico-Epidemiólogo

Tabla 4.18. Análisis de Formulas Estadísticas.

#### 4.4 Diagrama de Procedimiento Diseño de Formulas Estadísticas.

Durante el diseño de las formulas estadísticas, se establecerán las modificaciones detectadas como necesarias dentro de las formulas estadísticas seleccionadas para obtener los resultados esperados, la captura de los datos que estas fórmulas necesitan y también la forma en la que se presentarán dichos resultados.

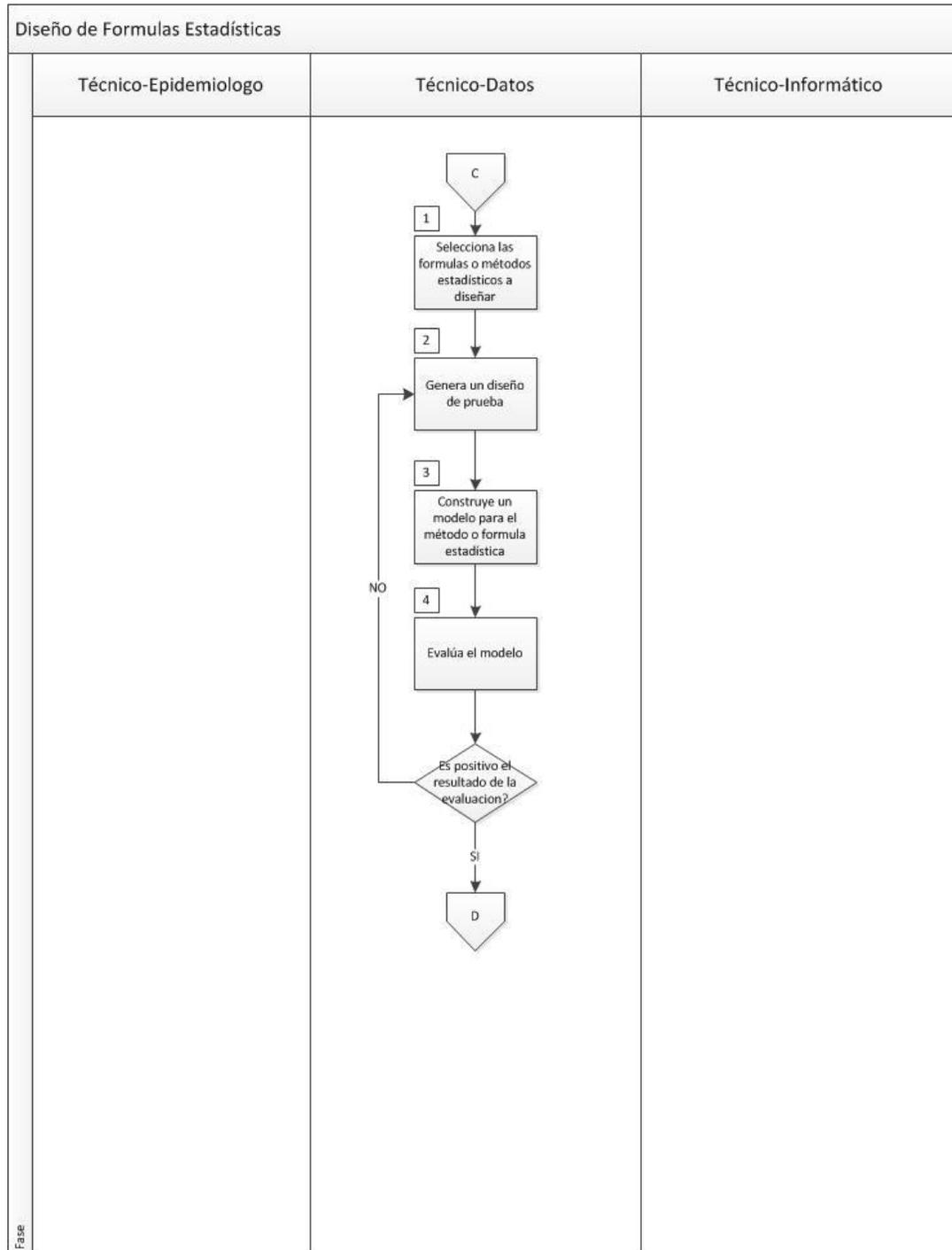


Fig.4.7- Diseño de Formulas Estadísticas.

#### 4.4.1 Descripción de Procedimiento Diseño de Formulas Estadísticas.

<b>Secuencia de Etapas</b>	<b>Actividad</b>	<b>Responsable</b>
1. Selecciona las formulas o métodos estadísticos a diseñar.	1.1 Define las funciones estadística as desarrollar.	Técnico-Datos
2. Genera un diseño de prueba.	2.1 Elaborar un diseño de prueba de los datos. 2.2 Decidir los pasos necesarios para la prueba. 2.3 Preparar los datos requeridos para la prueba.	Técnico-Datos
3. Construye un modelo para el método o formula estadística.	3.1 Determinar parámetros iniciales. 3.2 Ejecutar función o método seleccionado sobre un conjunto de datos de entrada para producir el modelo. 3.3 Elaborar una descripción detallada del modelo.	Técnico-Datos
4. Evalua el modelo	4.1 Evaluar los resultados obtenidos del modelado. 4.2 Probar los resultados obtenidos. 4.3 Seleccionar los mejores modelos. 4.4 Interpretar los resultados en términos del negocio. 4.5 Chequear la credibilidad del modelo.	Técnico-Datos

Tabla 4.19. Diseño de Formulas Estadísticas.

#### 4.5 Diagrama de Procedimiento Desarrollo de Formulas Estadísticas.

Partiendo del diseño obtenido en el paso anterior, se procederá a generar el conjunto de instrucciones necesarias para ejecutar las funciones estadísticas.

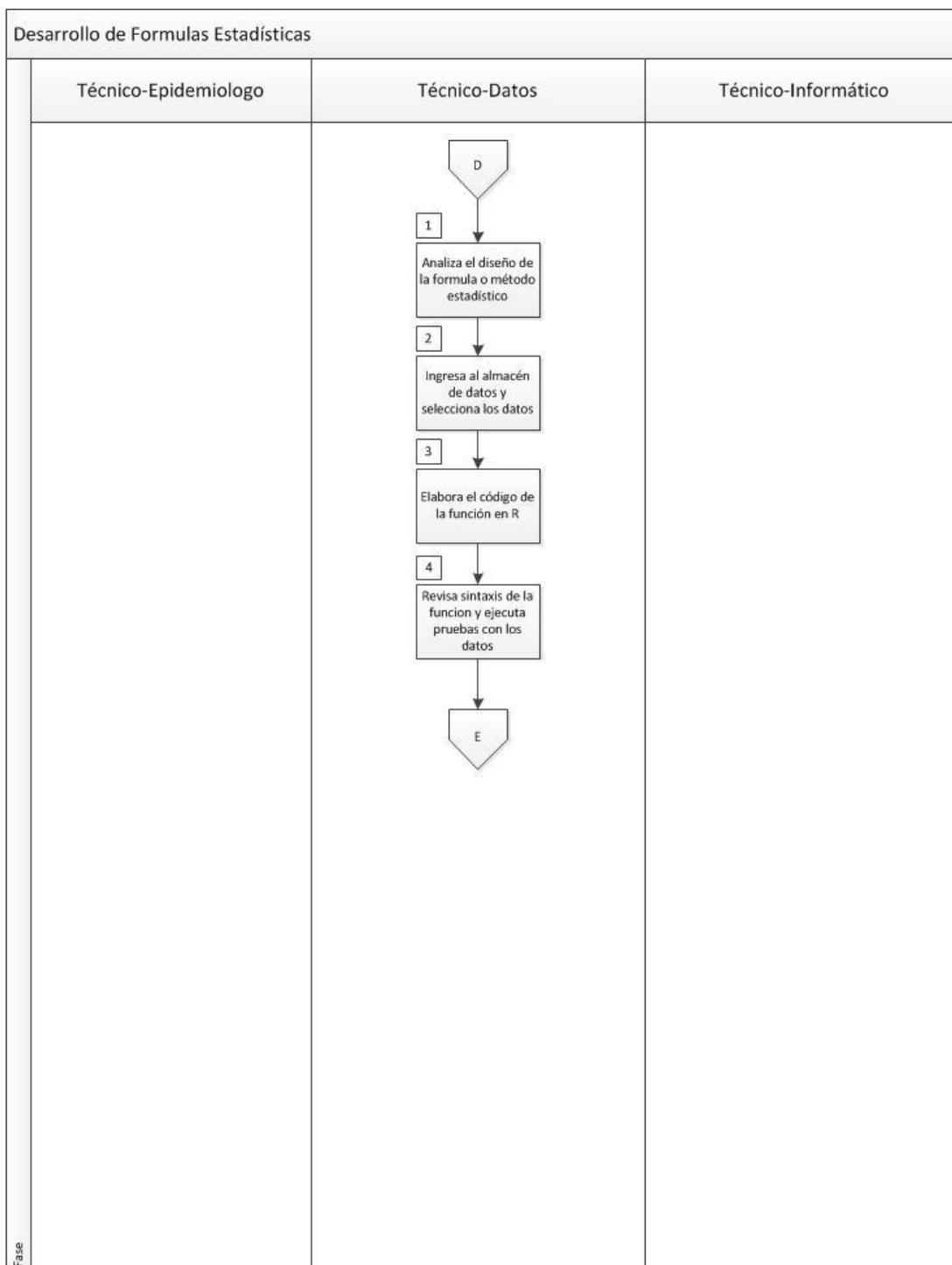


Fig.4.8-Desarrollo de Funciones Estadísticas.

#### 4.5.1 Descripción de Procedimiento Desarrollo de Formulas Estadísticas.

<b>Secuencia de Etapas</b>	<b>Actividad</b>	<b>Responsable</b>
1. Analiza el diseño de la formula o método estadístico.	1.1 Identifica Parámetros de entrada. 1.2 Identifica tablas y variables a utilizar. 1.3 Identifica restricciones en el diseño.	Técnico-Datos
2. Ingresa al almacén de datos y selecciona los datos.	2.1 Ingresa al almacén de datos. 2.2 Selecciona las tablas de datos a utilizar.	Técnico-Datos
3.Elabora el código de la función en R.	3.1 Crea la sintaxis de la función de acuerdo a los requerimientos del diseño.	Técnico-Datos
4. Revisa la sintaxis y ejecuta pruebas con los datos.	4.1 Verifica la sintaxis de la función. 4.2 Realiza las pruebas necesarias a la función.	Técnico-Datos

Tabla 4.20. Desarrollo de Formulas Estadísticas.

#### 4.6 Diagrama de Procedimiento Pruebas para Presentación de Resultados

Las formulas estadísticas deberán ser puestas a prueba de manera que generen los resultados esperados de forma correcta.

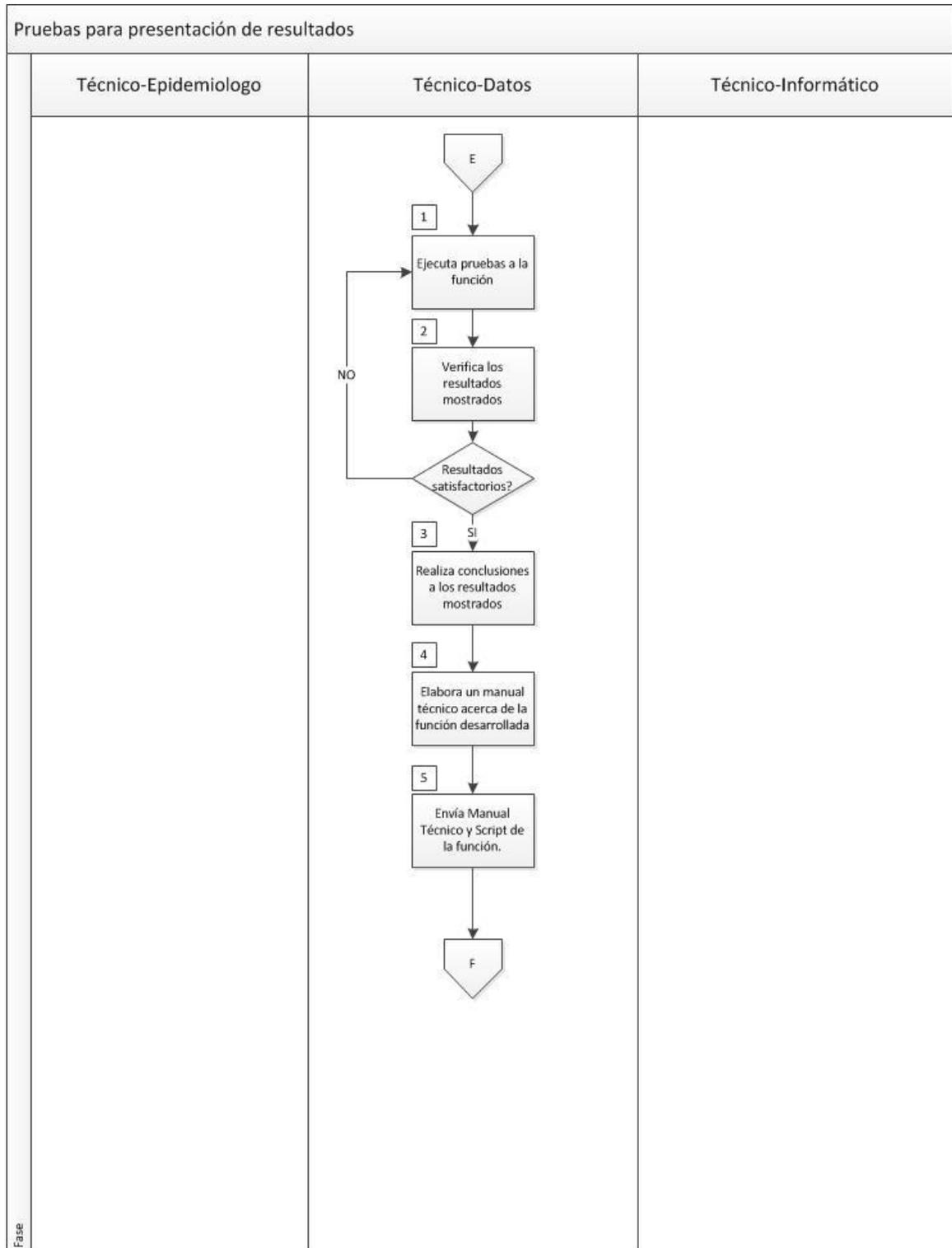


Fig.4.9- Pruebas para presentación de resultados.

#### 4.6.1 Descripción de Procedimiento Pruebas para Presentación de Resultados.

<b>Secuencia de Etapas</b>	<b>Actividad</b>	<b>Responsable</b>
1. Ejecuta pruebas a la función.	1.1 Realiza el número de pruebas necesarias hasta obtener el resultado esperado.	Técnico-Datos
2. Verifica los resultados mostrados.	2.1 Compara el resultado obtenido contra lo esperado según el diseño.	Técnico-Datos
3. Realiza conclusiones a los resultados mostrados.	3.1 Elabora conclusiones cuando el resultado es satisfactorio.	Técnico-Datos
4. Elabora un manual técnico acerca de la función desarrollada.	4.1 Elabora un manual técnico propio de la función desarrollada.	Técnico-Datos
5. Envía manual técnico y Script de la función.	5.1 Envía el manual técnico y la función desarrollada para su integración con SIIG.	Técnico-Datos

Tabla 4.21. Pruebas para Presentación de Resultados.

#### 4.7 Diagrama de Procedimiento Integración con el Modulo de Tablero Electrónico de Indicadores.

Una vez probadas las funciones estadísticas, deberán integrarse con el módulo de tablero de indicadores de salud para la presentación de sus resultados.

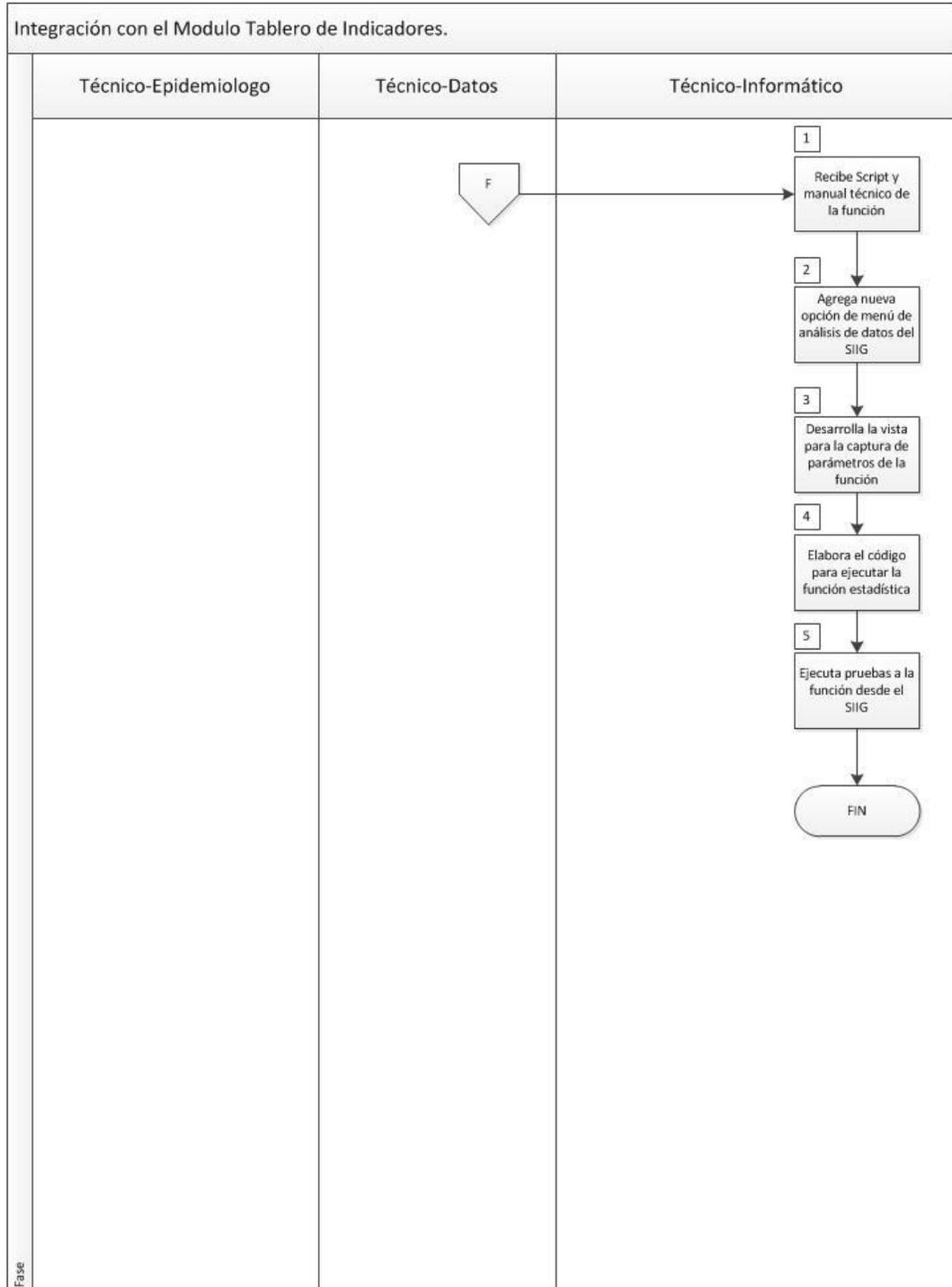


Fig.4.10- Integración con modulo tablero de indicadores.

#### 4.7.1 Descripción de Procedimiento Integración con el Modulo Tablero de Indicadores.

Secuencia de Etapas	Actividad	Responsable
1. Recibe Script y manual técnico de la función.	1.1 Verifica Script y manual técnico de la función.	Técnico-Informático
2. Agrega nueva opción de menú de análisis de datos del SIIG.	2.1 Dentro del menú análisis de datos desarrolla una nueva opción.	Técnico-Informático
3. Desarrolla la vista para la captura de parámetros de la función.	3.1 Desarrolla el código para la vista que se desplegara en pantalla.	Técnico-Informático
4. Elabora el código para llamar a la función estadística.	4.1 Elabora un modelo para ejecutar la función estadística. 4.2 Elabora el código para el controlador de la nueva opción de menú.	Técnico-Informático
5. Ejecuta pruebas a la función desde el SIIG.	5.1 Ejecuta las pruebas necesarias hasta que la ejecución de la función estadística esté libre de errores.	Técnico-Informático

Tabla 4.22. Integración de Funciones.

#### 5. Documentos de Referencia

- i. Programa “Construyendo La Esperanza – Estrategias y Recomendaciones en Salud 2009-2014”
- ii. Estructura Organizativa del Ministerio de Salud de El Salvador
- iii. Manual de Organización y Funciones de Las RIIS Enero 2012

---

---

# CAPÍTULO 5: DISEÑO

## **SINOPSIS**

En este capítulo se presenta un conjunto de bosquejos de pantallas de software y tablas con el diseño de las funciones estadísticas que permiten tener una visión preliminar del resultado final del desarrollo del proyecto.

---

---

## 5.1 DISEÑO DE LAS SALIDAS

### 5.1.1 MENÚ PRINCIPAL

Este menú tiene como función principal enlazar el acceso a los dos menús que permiten acceder a todas las funcionalidades que ofrece el Módulo de Análisis de Datos.



Análisis de Datos
Funciones Estadísticas
Tablero

Fig. 5.1 – Menú Principal

### 5.1.2 MENÚ FUNCIONES ESTADÍSTICAS

El menú funciones estadísticas tiene la finalidad de guardar información relativa a las funciones que están desarrolladas, por lo que permite agregar, modificar y eliminar datos. Tiene la siguiente estructura:



Listado de Funciones Estadísticas		
Nombre	Clasificación	Descripción

Fig. 5.2 – Menú Funciones Estadísticas

### 5.1.3 MENÚ TABLERO

El menú Tablero está compuesto por una serie de pestañas que permiten la generación y presentación de los resultados de las Funciones Estadísticas.



Fig. 5.3 – Menú Tablero

A continuación se presenta una descripción de cada una de las pestañas que conforman el menú Tablero:

- **Gráficos:** También conocida como la pestaña de resultados, en ésta se presentan los resultados de las funciones estadísticas, las salidas pueden ser graficas o cálculos numéricos.

#### VISTA TIPO TABLERO

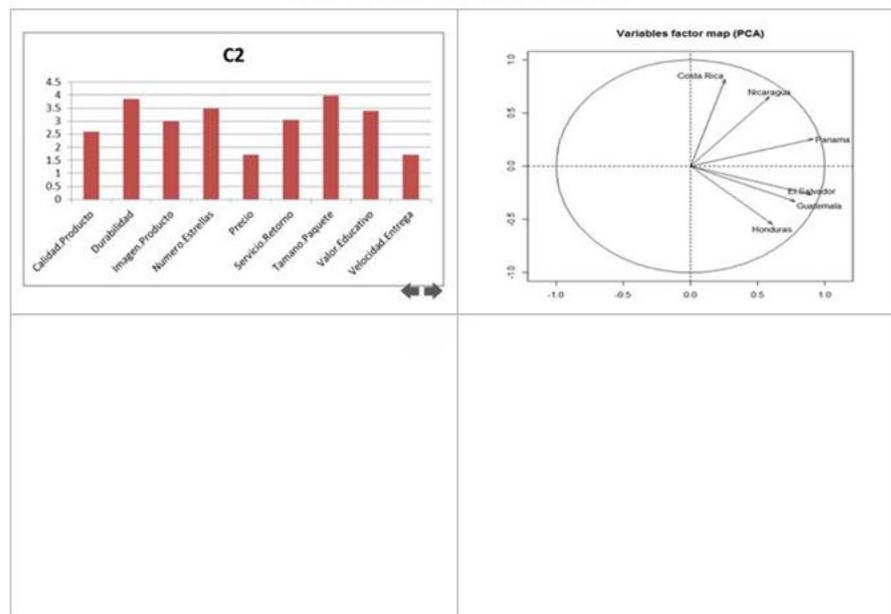


Fig. 5.4 – Pestaña Gráficos (también conocida como Resultados)

- **Salas:** Permite guardar los datos referente a los cálculos realizados por el usuario.



Fig. 5.5 – Pestaña Salas.

- **Indicadores:** Muestra el listado de indicadores que están disponibles en el SIIG y permite seleccionar un indicador y la función estadística a aplicar.

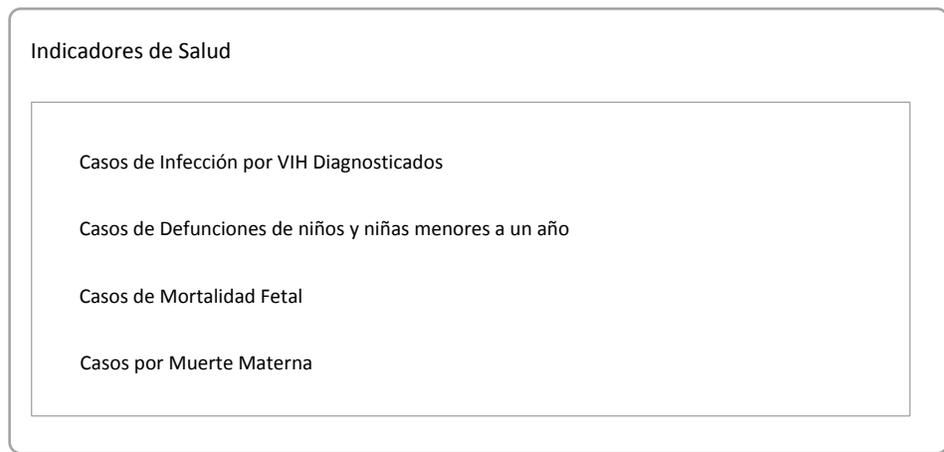


Fig. 5.6 – Pestaña Indicadores.

- **Análisis:** Permite al usuario poder escribir algún comentario o conclusión de los cálculos realizados.



Análisis

Escribir texto

Fig. 5.7 – Pestaña Análisis.

- **Filtros y Parámetros:** Presenta los formularios de cada función estadista de acuerdo a la seleccionada por el usuario



Filtros y Parametros

Filtrar Datos Por:

Desde Año:	Selec ▼	Hasta Año:	Selecci ▼
Desde Mes:	Selec ▼	Hasta Mes:	Selecci ▼
Region:	Selec ▼	Sexo:	Selecci ▼
Departamento:	Selec ▼	Municipio:	Selecci ▼
Area:	Selec ▼	Establecimiento:	Selecci ▼

Parámetros

**Resumen Estadístico**

Indicador Seleccionado: Índice de Mortalidad Fetal

Seleccione Variable de Indicador: Edad ▼

Aceptar Cancelar

Fig. 5.8 – Pestaña Filtros y Parámetros.

- **Ayuda:** Proporciona información al usuario acerca de la función que se desea utilizar.

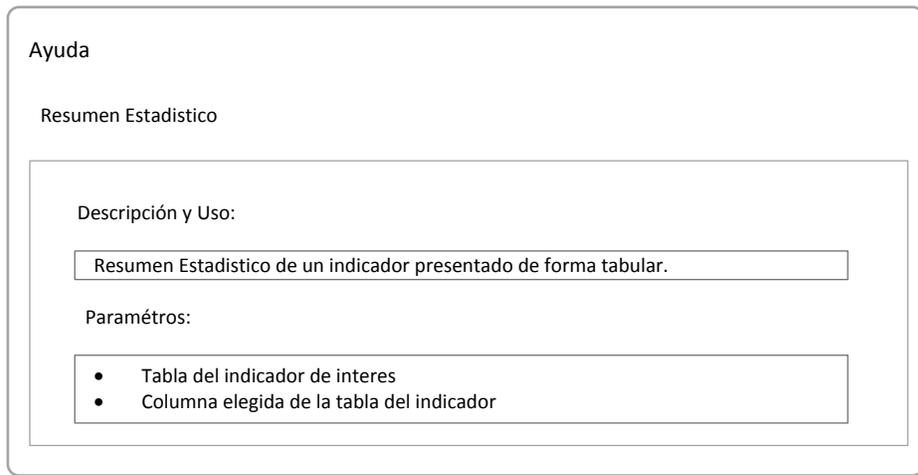


Fig. 5.9 – Pestaña Ayuda.

## 5.2 DISEÑO DE LAS FUNCIONES ESTADÍSTICAS

### I. RESUMEN DE MEDIDAS DE POSICIÓN, TENDENCIA CENTRAL Y DISPERSIÓN

<b>Nombre de la función:</b>	Resumen de medidas de posición, tendencia central y dispersión.					
<b>Objetivo:</b>	Presentar de forma resumida los valores de las medidas de posición, tendencia central y dispersión de acuerdo con los valores que generan un indicador determinado.					
<b>Nombre del indicador que aplica:</b>	Todos los indicadores de manera individual					
<b>Fuente de datos:</b>	La tabla de valores a partir de los cuales se crea cada indicador					
<b>Datos de entrada:</b>						
	Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo
	Tabla del Indicador		x	x	Alfanumérico	Varchar 25
	Variable del Indicador		x	x	Alfanumérico	Varchar 25
<b>Procesos involucrados en la función:</b>	Carga de Datos Categorización de Datos Filtrado de Datos Presentación de Datos					
<b>Filtros:</b>	Por año y/o mes Por Región Por departamento o municipio Por sexo Por establecimiento Por área					
<b>Frecuencia de uso:</b>	Mensualmente					

**Pantalla de Entrada:**

**Resumen Estadístico**

Indicador Seleccionado:

Selecione Variable de Indicador:

Fig. 5.10 – Pantalla de entrada de la función Resumen Estadístico.

**Nombre código de las función:**

`_plr_resumenestadistico`

Tabla 5.1. Resumen Estadístico.

## II. REGRESIÓN LINEAL SIMPLE

<b>Nombre de la función:</b>	Regresión Lineal Simple				
<b>Objetivo:</b>	Desarrollar el medio por el cual se pueda verificar si los datos de un indicador están relacionados de forma lineal.				
<b>Nombre del indicador que aplica:</b>	Todos los indicadores de manera individual				
<b>Fuente de datos:</b>	La tabla de valores a partir de los cuales se crea cada indicador				
<b>Datos de entrada:</b>					
Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo
Tabla del Indicador		x	x	Alfanumérico	Varchar 25
Variable x		x	x	Alfanumérico	Varchar 25
Variable y		x	x	Alfanumérico	Varchar 25
Título del Grafico	x			Alfanumérico	Varchar 50
Título del eje x	x			Alfanumérico	Varchar 25
Título del eje y	x			Alfanumérico	Varchar 25
<b>Procesos involucrados en la función:</b>	Carga de Datos Categorización de Datos Filtrado de Datos Presentación de Datos				
<b>Filtros:</b>	Por año y/o mes Por Región Por departamento o municipio Por sexo Por establecimiento Por área				
<b>Frecuencia de uso:</b>	Mensualmente				

**Pantalla de Entrada:**

**Regresión Lineal**

Indicador Seleccionado:

Variable x:

Variable y:

Personalización del Gráfico:

Título del gráfico:

Título del eje x:

Título del eje y:

Fig. 5.11 – Pantalla de entrada de la función Regresión Lineal

<b>Nombre código de las función:</b>	<code>_plr_regresionlineal</code>
--------------------------------------	-----------------------------------

Tabla 5.2. Regresión Lineal.

### III. INTERVALO DE CONFIANZA

<b>Nombre de la función:</b>	Intervalo de Confianza				
<b>Objetivo:</b>	Desarrollar una función que permita encontrar un intervalo de confianza a un conjunto de datos de una variable cuantitativa.				
<b>Nombre del indicador que aplica:</b>	Todos los indicadores de manera individual				
<b>Fuente de datos:</b>	La tabla de valores a partir de los cuales se crea cada indicador				
<b>Datos de entrada:</b>					
Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo
Tabla del Indicador		x	x	Alfanumérico	Varchar 25
Variable Indicador		x	x	Alfanumérico	Varchar 25
Nivel de Confianza		x	x	Decimal	Double
<b>Procesos involucrados en la función:</b>	Carga de Datos Categorización de Datos Filtrado de Datos Presentación de Datos				
<b>Filtros:</b>	Por año y/o mes Por región Por departamento o municipio Por sexo Por establecimiento Por área				
<b>Frecuencia de uso:</b>	Mensualmente				

**Pantallas de Entrada:**

**Intervalo de Confianza**

Indicador Seleccionado:

Nombre Variable Indicador:

Nivel de Confianza:

Fig. 5.12 – Pantalla de entrada de la función Intervalo de Confianza.

**Nombre código de las función:**

`_plr_intervaloconf`

Tabla 5.3. Métodos Inferenciales.

#### IV. MATRIZ DE CORRELACIÓN

<b>Nombre de la función:</b>	Matriz de Correlación												
<b>Objetivo:</b>	Desarrollar el medio por el cual se pueda calcular la matriz de correlaciones de los indicadores de salud.												
<b>Nombre del indicador que aplica:</b>	Todos los indicadores de manera individual												
<b>Fuente de datos:</b>	La tabla de valores a partir de los cuales se crea cada indicador												
<b>Datos de entrada:</b>													
<table border="1"> <thead> <tr> <th>Elemento</th> <th>Digitado</th> <th>Seleccionado</th> <th>Recuperado</th> <th>Formato</th> <th>Tipo</th> </tr> </thead> <tbody> <tr> <td>Tabla del Indicador</td> <td></td> <td>x</td> <td>x</td> <td>Alfanumérico</td> <td>Varchar 25</td> </tr> </tbody> </table>	Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo	Tabla del Indicador		x	x	Alfanumérico	Varchar 25	
Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo								
Tabla del Indicador		x	x	Alfanumérico	Varchar 25								
<b>Procesos involucrados en la función:</b>	Carga de Datos Categorización de Datos Filtrado de Datos Presentación de Datos												
<b>Filtros:</b>	Por año y/o mes Por Región Por departamento o municipio Por sexo Por establecimiento Por área												
<b>Frecuencia de uso:</b>	Mensualmente												

**Pantalla de Entrada:**

**Matriz de Correlación**

Indicador Seleccionado:

Fig. 5.13 – Pantalla de entrada de la función Matriz de Correlación.

**Nombre código de las función:**

\_plr\_matrizcorr

Tabla 5.4. Matriz de Correlación

## V. MAPA DE CORRELACIÓN

<b>Nombre de la función:</b>	Mapa de Correlación				
<b>Objetivo:</b>	Desarrollar el medio por el cual se pueda generar el mapa de correlaciones de los datos de los indicadores de salud.				
<b>Nombre del indicador que aplica:</b>	Todos los indicadores de manera individual				
<b>Fuente de datos:</b>	La tabla de valores a partir de los cuales se crea cada indicador.				
<b>Datos de entrada:</b>					
Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo
Tabla del Indicador		x	x	Alfanumérico	Varchar 25
Método		x		Booleano	Booleano
Título del grafico	x			Alfanumérico	Varchar 50
Título del eje x	x			Alfanumérico	Varchar 25
Título del eje y	x			Alfanumérico	Varchar 25
Escala Valores		x		Booleano	Booleano
<b>Procesos involucrados en la función:</b>	Carga de Datos Categorización de Datos Filtrado de Datos Presentación de Datos				
<b>Filtros:</b>	Por año y/o mes Por Región Por departamento o municipio Por sexo Por establecimiento Por área				
<b>Frecuencia de uso:</b>	Mensualmente				

**Pantalla de Entrada:**

**Mapa de Correlaciones**

Indicador Seleccionado:

Método:  Pearson  Kendall  Spearman

Personalización del Gráfico:

Título del gráfico:

Título del eje x:

Título del eje y:

Mostrar Escala de Valores:

Fig. 5.14 – Pantalla de entrada de la función Mapa de Correlación.

**Nombre código de las función:**

`_plr_matrizcorr`

Tabla 5.5. Mapa de Correlación

## VI. CÍRCULO DE CORRELACIÓN

<b>Nombre de la función:</b>	Circulo de Correlación				
<b>Objetivo:</b>	Desarrollar el medio por el cual se pueda generar el círculo de correlación de los datos de los indicadores de salud.				
<b>Nombre del indicador que aplica:</b>	Todos los indicadores de manera individual				
<b>Fuente de datos:</b>	La tabla de valores a partir de los cuales se crea cada indicador				
<b>Datos de entrada:</b>					
Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo
Tabla del Indicador		x	x	Alfanumérico	Varchar 25
Número de Componentes Principales		x		Numérico	Integer
<b>Procesos involucrados en la función:</b>	Carga de Datos Categorización de Datos Filtrado de Datos Presentación de Datos				
<b>Filtros:</b>	Por año y/o mes Por Región Por departamento o municipio Por sexo Por establecimiento Por área				
<b>Frecuencia de uso:</b>	Mensualmente				

**Pantalla de Entrada:**

**Circulo de Correlaciones**

Indicador Seleccionado:

Número de Componentes Principales:

Fig. 5.15 – Pantalla de entrada de la función Circulo de Correlación.

<b>Nombre código de las función:</b>	<code>_plr_circulocorr</code>
--------------------------------------	-------------------------------

Tabla 5.6. Circulo de Correlaciones

## VII. CLÚSTERES EN EL PLANO/DENDOGRAMA

<b>Nombre de la función:</b>	Clústeres en el plano/dendograma				
<b>Objetivo:</b>	Desarrollar el medio por el cual se pueda generar clústeres en el plano y dendogramas de las variables que generan a los indicadores de salud.				
<b>Nombre del indicador que aplica:</b>	Todos los indicadores de manera individual				
<b>Fuente de datos:</b>	La tabla de valores a partir de los cuales se crea cada indicador				
<b>Datos de entrada:</b>					
Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo
Tabla del Indicador		x	x	Alfanumérico	Varchar 25
Método		x		Booleano	Booleano
Cantidad de Clúster		x		Numérico	Integer
<b>Procesos involucrados en la función:</b>	Carga de Datos Categorización de Datos Filtrado de Datos Presentación de Datos				
<b>Filtros:</b>	Por año y/o mes Por Región Por departamento o municipio Por sexo Por establecimiento Por área				
<b>Frecuencia de uso:</b>	Mensualmente				

**Pantalla de Entrada:**

**Dendograma**

Indicador Seleccionado:

Método:  Complete  Average  Single  Ward

Cantidad de Cluster:

Fig. 5.16 – Pantalla de entrada de la función Dendograma.

**Nombre código de las función:**

\_plr\_dendograma

Tabla 5.7. Dendograma.

## VIII. ANÁLISIS DE CONGLOMERADOS

<b>Nombre de la función:</b>	Análisis de Conglomerados					
<b>Objetivo:</b>	Desarrollar el medio por el cual se pueda generar el análisis de conglomerado, también conocido como clústeres en el plano de las variables que generan a los indicadores de salud.					
<b>Nombre del indicador que aplica:</b>	Todos los indicadores de manera individual					
<b>Fuente de datos:</b>	La tabla de valores a partir de los cuales se crea cada indicador					
<b>Datos de entrada:</b>						
	Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo
	Tabla del Indicador		x	x	Alfanumérico	Varchar 25
	Método		x		Booleano	Booleano
	Cantidad de Clúster		x		Numérico	Integer
<b>Procesos involucrados en la función:</b>	Carga de Datos Categorización de Datos Filtrado de Datos Presentación de Datos					
<b>Filtros:</b>	Por año y/o mes Por Región Por departamento o municipio Por sexo Por establecimiento Por área					
<b>Frecuencia de uso:</b>	Mensualmente					

**Pantalla de Entrada:**

**Análisis de Conglomerados**

Indicador Seleccionado:

Número de Clusters:

Personalización del Gráfico:

Título del gráfico:

Título del eje x:

Título del eje y:

Enumerar Conglomerados:

Fig. 5.17 – Pantalla de entrada de la función Análisis de Conglomerados.

**Nombre código de las función:**

\_plr\_clusterizacion

Tabla 5.8. Análisis de Conglomerados

IX. DIAGRAMA DE CAJA Y BIGOTE

<b>Nombre de la función:</b>	Diagrama de Caja y Bigote				
<b>Objetivo:</b>	Desarrollar el medio por el cual se pueda generar el diagrama de caja y bigote de las variables que generan a los indicadores de salud.				
<b>Nombre del indicador que aplica:</b>	Todos los indicadores de manera individual				
<b>Fuente de datos:</b>	La tabla de valores a partir de los cuales se crea cada indicador.				
<b>Datos de entrada:</b>					
Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo
Tabla del Indicador		x	x	Alfanumérico	Varchar 25
Mostrar atípicos		x		Booleano	Booleano
Título del grafico	x			Alfanumérico	Varchar 50
Título del eje x	x			Alfanumérico	Varchar 25
Título del eje y	x			Alfanumérico	Varchar 25
<b>Procesos involucrados en la función:</b>	Carga de Datos Categorización de Datos Filtrado de Datos Presentación de Datos				
<b>Filtros:</b>	Por año y/o mes Por Región Por departamento o municipio Por sexo Por establecimiento Por área				
<b>Frecuencia de uso:</b>	Mensualmente				

**Pantalla de Entrada:**

**Diagrama Caja y Bigote**

Indicador Seleccionado:

Mostrar Atípicos:

Personalización del Gráfico:

Título del gráfico:

Título del eje x:

Título del eje y:

Fig. 5.18 – Pantalla de entrada de la función Diagrama de Caja y Bigote.

**Nombre código de las función:**

`_plr_cajabigote`

Tabla 5.9. Diagrama de Caja y Bigote

## X. CENTROS DE GRAVEDAD

<b>Nombre de la función:</b>	Centros de gravedad				
<b>Objetivo:</b>	Desarrollar el medio por el cual se pueda calcular los centros de gravedad de los indicadores de salud.				
<b>Nombre del indicador que aplica:</b>	Todos los indicadores de manera individual				
<b>Fuente de datos:</b>	La tabla de valores a partir de los cuales se crea cada indicador				
<b>Datos de entrada:</b>					
Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo
Tabla del Indicador		x	x	Alfanumérico	Varchar 25
Método		x		Booleano	Booleano
Resultado		x		Booleano	Booleano
<b>Procesos involucrados en la función:</b>	Carga de Datos Categorización de Datos Filtrado de Datos Presentación de Datos				
<b>Filtros:</b>	Por año y/o mes Por Región Por departamento o municipio Por sexo Por establecimiento Por área				
<b>Frecuencia de uso:</b>	Mensualmente				

**Pantalla de Entrada:**

**Centros de Gravedad**

Indicador Seleccionado:

Método:  Complete  Average  Single  Ward

Resultado como:  Histograma  Calculo Numérico

Fig. 5.19 – Pantalla de entrada de la función Centros de Gravedad.

**Nombre código de las función:**

\_plr\_centrosg

Tabla 5.10. Centros de Gravedad.

## XI. DATOS ATÍPICOS O MAL REPRESENTADOS

<b>Nombre de la función:</b>	Datos atípicos o mal representados																		
<b>Objetivo:</b>	Desarrollar el medio por el cual se identifiquen los datos atípicos o mal representados de los valores que generan a los indicadores de salud.																		
<b>Nombre del indicador que aplica:</b>	Todos los indicadores de manera individual																		
<b>Fuente de datos:</b>	La tabla de valores a partir de los cuales se crea cada indicador																		
<b>Datos de entrada:</b>																			
<table border="1"> <thead> <tr> <th>Elemento</th> <th>Digitado</th> <th>Seleccionado</th> <th>Recuperado</th> <th>Formato</th> <th>Tipo</th> </tr> </thead> <tbody> <tr> <td>Tabla del Indicador</td> <td></td> <td>x</td> <td>x</td> <td>Alfanumérico</td> <td>Varchar 25</td> </tr> <tr> <td>Datos a Mostrar</td> <td></td> <td>x</td> <td></td> <td>Numérico</td> <td>Integer</td> </tr> </tbody> </table>	Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo	Tabla del Indicador		x	x	Alfanumérico	Varchar 25	Datos a Mostrar		x		Numérico	Integer	
Elemento	Digitado	Seleccionado	Recuperado	Formato	Tipo														
Tabla del Indicador		x	x	Alfanumérico	Varchar 25														
Datos a Mostrar		x		Numérico	Integer														
<b>Procesos involucrados en la función:</b>	Carga de Datos Categorización de Datos Filtrado de Datos Presentación de Datos																		
<b>Filtros:</b>	Por año y/o mes Por Región Por departamento o municipio Por sexo Por establecimiento Por área																		
<b>Frecuencia de uso:</b>	Mensualmente																		

**Pantalla de Entrada:**

**Datos Atípicos**

Indicador Seleccionado:

Datos a Mostrar:

Fig. 5.20 – Pantalla de entrada de la función Datos Atípicos.

**Nombre código de las función:**

`_plr_atipicos`

Tabla 5.11. Datos Atípicos

## 5.3 DISEÑO DE LOS PROCESOS DEL MÓDULO DE ANÁLISIS DE DATOS

### A. Salida: Matriz de Correlación

#### Proceso Lógico

fechaInic ← Fecha inicial

fechaFin ← Fecha Final

nombreIndicador ← Tabla de indicador

variablesSeleccionadas ← Variables del indicador seleccionadas

consultaSQL ← SELECT variablesSeleccionadas FROM nombreIndicador i WHERE i.fecha BETWEEN fechaInic AND fechaFin

Si (((fechaInic && fechaFin) <> null) && (fechaInic <= fechaFin))

{

**Ejecutar** consulta SQL

**Guardar** datos en tabla temporal

**Definir** parámetros para llamar funciones de R.

**Llamar** funciones de R para calcular la matriz de correlación

**Obtener** la matriz de correlación

}

**Mostrar** en pantalla la matriz de correlación

### B. Salida: Mapa de Correlación

#### Proceso Lógico

fechaInic ← Fecha inicial

fechaFin ← Fecha Final

nombreIndicador ← Tabla de indicador

variablesSeleccionadas ← Variables del indicador seleccionadas

consultaSQL ← SELECT variablesSeleccionadas FROM nombreIndicador i WHERE i.fecha BETWEEN fechaInic AND fechaFin

Si (((fechaInic && fechaFin) <> null) && (fechaInic <= fechaFin))

{

**Ejecutar** consulta SQL

**Guardar** datos en tabla temporal

**Definir** parámetros para llamar funciones de R.

**Llamar** funciones de R para calcular el mapa de correlación

**Obtener** matriz de correlación

}

**Mostrar** en pantalla el mapa de correlación

### C. Salida: Círculo de Correlación

#### Proceso Lógico

fechaInic ← Fecha inicial

fechaFin ← Fecha Final

nombreIndicador ← Tabla de indicador

variablesSeleccionadas ← Variables del indicador seleccionadas

consultaSQL ← SELECT variablesSeleccionadas FROM nombreIndicador i WHERE i.fecha BETWEEN fechaInic AND fechaFin

Si (((fechaInic && fechaFin) <> null) && (fechaInic <= fechaFin))

{

**Ejecutar** consulta SQL

**Guardar** datos en tabla temporal

**Definir** parámetros para llamar funciones de R.

**Llamar** funciones de R para calcular el círculo de correlación.

**Obtener** matriz de correlación

}

**Mostrar** en pantalla el círculo de correlación

### D. Salida: Clústeres en el plano/dendograma

#### Proceso Lógico

fechaInic ← Fecha inicial

fechaFin ← Fecha Final

nombreIndicador ← Tabla de indicador

cantCluster ← Cantidad de clústeres

tipoGrafico ← Tipo de grafico

variablesSeleccionadas← Variables del indicador seleccionadas

consultaSQL← SELECT variablesSeleccionadas FROM nombreIndicador i WHERE i.fecha BETWEEN fechInic AND fechFin

Si (((fechInic&&fechFin)<>null)&&(fechInic<=fechFin)) && cantCluster>=1)

{

**Ejecutar** consulta SQL

**Guardar** datos en tabla temporal

Si (tipoGrafico=0){

**Definir** parámetros para llamar funciones de R.

**Llamar** funciones de R para calcular el plano.

**Obtener** el plano

}

**Mostrar** en pantalla los datos en el plano

Si (tipoGrafico=1){

**Definir** parámetros para llamar funciones de R.

**Llamar** funciones de R para calcular el dendograma.

**Obtener** el dendograma

}

**Mostrar** en pantalla el dendograma

}

## E. Salida: Centros de Gravedad

### Proceso Lógico

fechInic←Fecha inicial

fechFin← Fecha Final

nombreIndicador← Tabla de indicador

cantCluster ←Cantidad de clústeres

variablesSeleccionadas← Variables del indicador seleccionadas

consultaSQL← SELECT variablesSeleccionadas FROM nombreIndicador i WHERE i.fecha BETWEEN fechInic AND fechFin

```

Si (((fechInic&&fechFin)<>null)&&(fechInic<=fechFin)) && cantCluster>=1)
{
    Ejecutar consulta SQL
    Guardar datos en tabla temporal
    Definir parámetros para llamar funciones de R.
    Llamar funciones de R para calcular los centros de gravedad.
    Obtener centros de gravedad
}

```

**Mostrar** en pantalla los centros de gravedad

## F. Salida: Gráficos Horizontales o Verticales para los centros de gravedad

### Proceso Lógico

fechInic←Fecha inicial

fechFin← Fecha Final

nombreIndicador← Tabla de indicador

cantCluster ←Cantidad de clústeres

variablesSeleccionadas← Variables del indicador seleccionadas

consultaSQL← SELECT variablesSeleccionadas FROM nombreIndicador i WHERE i.fecha BETWEEN fechInic AND fechFin

```

Si (((fechInic&&fechFin)<>null)&&(fechInic<=fechFin)) && cantCluster>=1)

```

```

{
    Ejecutar consulta SQL
    Guardar datos en tabla temporal
    Si (tipoGrafico=0){
        Definir parámetros para llamar funciones de R.
        Llamar funciones de R para calcular grafico vertical
        Obtener el grafico vertical de los centros de gravedad
    }
}

```

**Mostrar** en pantalla el grafico vertical

```

Si (tipoGrafico=1){

```

```

    Definir parámetros para llamar funciones de R.

```

**Llamar** funciones de R para calcular grafico horizontal  
**Obtener** el grafico horizontal de los centros de gravedad  
}

**Mostrar** en pantalla el grafico horizontal  
}

## G. Salida: Histograma representativo para un indicador de salud

### Proceso Lógico

fechaInic ← Fecha inicial

fechaFin ← Fecha Final

nombreIndicador ← Tabla de indicador

variablesSeleccionadas ← Variables del indicador seleccionadas

consultaSQL ← SELECT variablesSeleccionadas FROM nombreIndicador i WHERE i.fecha  
BETWEEN fechaInic AND fechaFin

Si (((fechaInic && fechaFin) <> null) && (fechaInic <= fechaFin))

{  
**Ejecutar** consulta SQL  
**Guardar** datos en tabla temporal  
**Definir** parámetros para llamar funciones de R.  
**Llamar** funciones de R para calcular los valores del histograma.  
**Obtener** el histograma de los datos del indicador  
}

**Mostrar** en pantalla el histograma del indicador

## H. Salida: Diagrama de caja y bigote

### Proceso Lógico

fechaInic ← Fecha inicial

fechaFin ← Fecha Final

nombreIndicador ← Tabla de indicador

variablesSeleccionadas ← Variables del indicador seleccionadas

consultaSQL ← SELECT variablesSeleccionadas FROM nombreIndicador i WHERE i.fecha  
BETWEEN fechaInic AND fechaFin

Si (((fechInic&&fechFin)<>null)&&(fechInic<=fechFin)))

{

**Ejecutar** consulta SQL

**Guardar** datos en tabla temporal

**Definir** parámetros para llamar funciones de R.

**Llamar** funciones de R para calcular los valores del diagrama de caja y bigote.

**Obtener** el diagrama de caja y bigote de los datos del indicador

}

**Mostrar** en pantalla el gráfico de caja y bigote del indicador

## I. Salida: Datos atípicos o mal representados

### Proceso Lógico

fechInic←Fecha inicial

fechFin← Fecha Final

nombreIndicador← Tabla de indicador

variablesSeleccionadas← Variables del indicador seleccionadas

consultaSQL← SELECT variablesSeleccionadas FROM nombreIndicador i WHERE i.fecha BETWEEN fechInic AND fechFin

Si (((fechInic&&fechFin)<>null)&&(fechInic<=fechFin)))

{

Ejecutar consultaSQL

Guardar datos en tabla temporal

Definir parámetros para llamar funciones de R.

Llamar funciones de R para calcular los valores atípicos del indicador seleccionado

Obtener los valores atípicos de los datos del indicador

}

**Mostrar** en pantalla los valores atípicos del indicador

## J. Salida: Resumen de medidas estadísticas para un indicador

### Proceso Lógico

fechInic←Fecha inicial

fechaFin ← Fecha Final

nombreIndicador ← Tabla de indicador

variablesSeleccionadas ← Variables del indicador seleccionadas

consultaSQL ← SELECT variablesSeleccionadas FROM nombreIndicador i WHERE i.fecha BETWEEN fechaInic AND fechaFin

Si (((fechaInic && fechaFin) <> null) && (fechaInic <= fechaFin))

{

**Ejecutar** consulta SQL

**Guardar** datos en tabla temporal

**Definir** parámetros para llamar funciones de R.

**Llamar** funciones de R para calcular el resumen de medidas estadísticas del indicador seleccionado

**Obtener** el resumen de medidas estadísticas de los datos del indicador

}

**Mostrar** en pantalla el resumen de medidas estadísticas de los datos del indicador

## 5.4 DISEÑO DE LAS INTERFASES GRÁFICAS DE USUARIO

Para el desarrollo de la herramienta se tomara de base la interfaz gráfica del SIIG/ETAB, esto porque el análisis de datos se integrara como parte de este sistema.

### 5.4.1 INTERFAZ GENERAL DEL SIIG/ETAB

The screenshot shows the main interface of the SIIG/ETAB system. At the top, there is a header with the logo of the Ministry of Health of El Salvador, the text 'salud.gob.sv', and the user name 'admin'. Below the header is a navigation menu with options: 'Catálogos', 'Orígenes de datos', 'Indicadores', and 'Usuarios y grupos'. A search bar is located below the menu, and a 'Agregar nuevo' button is on the right. The main content area features a table with the following data:

Contacto	Nombre del establecimiento	Correo Electrónico	Número telefónico	Cargo
nombre persona contacto ejemplo 1	Establecimiento-lugar fuente de datos, ejemplo1	correo persona contacto ejemplo1	9999-9999	cargo persona contacto ejemplo 1

Below the table, there is an 'Exportar' button with options for 'json', 'xml', 'csv', and 'xls', and a 'Por página' dropdown menu set to '25'. On the right side, there is a 'Filtros' sidebar with input fields for 'Contacto' and 'Establecimiento', and 'Filtrar' and 'Resetear' buttons.

Fig. 5.21- Interfaz principal SIIG/ETAB.

#### Descripción de los elementos

1. Nombre del usuario actual y la opción para salir.
2. Logo/Imagen principal
3. Menú principal, se mostrarán las opciones de acuerdo al perfil del usuario.  
En el menú principal se agregara la opción de análisis de los indicadores, que mostrara todas las técnicas descriptivas disponibles para esta herramienta.
4. Barra de navegación. Contendrá las opciones en las cuales se ha ingresado
5. Área de trabajo

---

---

# **CAPÍTULO 6: DESARROLLO DEL MÓDULO DE ANÁLISIS DE DATOS E IMPLEMENTACIÓN**

## **SINOPSIS**

En este capítulo se presentan los estándares seguidos durante el desarrollo del proyecto y la forma para poder implementarlo dentro del sistema actual.

---

---

## 6.1 ESTÁNDARES DE DESARROLLO

### 6.1.1 ESTÁNDARES DE LA BASE DE DATOS

De acuerdo a los siguientes prefijos se pueden identificar los elementos creados en la base de datos del SIIG.

ELEMENTO	PREFIJO
Tablas catálogos	ctl
Scripts Funciones Estadísticas	_plr_
Tablas temporales de almacén de datos	tmp_ind_
Tablas temporales para análisis de datos	add_tmp_ind_

Tabla 6.1.Estandares Base de Datos

#### Estándares para asignar los nombres a los elementos de la base de datos

Para los nombres de los elementos de la base de datos se definieron las siguientes normas:

- La cantidad máxima de letras en los nombres de tablas incluyendo las temporales es de 63.
- Los nombres de los elementos siempre son con letras minúsculas.
- No deben existir espacios en blanco entre palabras, se utiliza el guion bajo (\_) para separar las palabras.

#### Ejemplos de nombres de elementos en la base de datos:

Descripción del Elemento	Prefijo	Nombre del Elemento	Nombre establecido según las normas
Tabla de departamentos	ctl	departamento	ctldepartamento
Función estadística para graficar un dendograma	_plr_	Dendograma	_plr_dendograma
Tabla del indicador número de mortalidad fetal	tmp_ind_	Numero de mortalidad fetal	tmp_ind_numero_de_mortalidad_fetal
Tabla filtrada del indicador número de mortalidad fetal	add_tmp_ind_	Numero de mortalidad fetal	add_tmp_ind_numero_de_mortalidad_fetal

Tabla 6.2.Estandares Tablas

## **Normas para asignar el nombre de los elementos de los campos de las tablas:**

- La cantidad de letras para el nombre de un campo de una tabla es de máximo 30 caracteres.
- Para separar las palabras en el nombre de un campo se utiliza el guion bajo(\_).
- De acuerdo a los estándares de symfony2 todas las tablas tendrán una llave primaria de nombre id.

## **6.1.2 ESTÁNDARES DE PROGRAMACIÓN**

Para el desarrollo del código del módulo de análisis de datos se utilizaron los estándares del Framework Symfony.

Cuando se programa en php conviene mantener la coherencia siguiendo unas convenciones en el código que facilitan la lectura. Al utilizar las mismas convenciones en el código es más fácil para nosotros y para otros programadores poder utilizar el mismo código.

A continuación un resumen de los estándares de Symfony.

### **General**

- Los ficheros deben usar solo las etiquetas de apertura php `<?php` o `<?='`
- Los ficheros deben usar solo UTF-8 sin BOM en el código php. El BOM es un residuo que permite a algunos editores de texto identificar la codificación.
- Los ficheros deben usar el formato de fichero Unix LF (lineFeed) en vez del usado en sistemas DOS (carryReturn).
- Los ficheros deberían o declarar símbolos (clases, funciones, constantes etc) o producir efectos (cómo generar salida, cambiar configuraciones, etc.) pero no ambos a la vez.
- Los nombres de clases deben ser declarados con estilo StudlyCaps.
- La declaración de las constantes en una clase debe usar letras mayúsculas con subrayado como separador.
- Los nombres de los métodos deben ser declarados en camelCase.

### **Estructura**

- El código debe usar cuatro espacios para la indentación en vez de usar el tabulado. Esto minimiza problemas con otras herramientas de desarrollo.
- Las líneas podrían tener 80 caracteres o menos evitando tener más de 120 caracteres.

- Las llaves de apertura deben ir en la siguiente línea y la llave de cierre debe ir en la siguiente línea después del cuerpo.
- Las llaves de apertura en las estructuras de control debe ir en la misma línea y las llaves de cierre deben de ir después del cuerpo.
- Los paréntesis en las estructuras de control no deben usar espacios antes o después.
- Añadir un solo espacio después de cada limitador de coma.
- Añadir un solo espacio alrededor de los operadores (==, &&, ...).
- Añadir una línea en blanco antes de una declaración de return, a no ser que esté dentro de una declaración como un grupo como if.
- Usa llaves para indicar el control de la estructura sin tener en cuenta el número de declaraciones que el grupo pueda contener.
- Define una clase por fichero.
- Declara las propiedades de clase antes que los propios métodos de clase.
- Declara los métodos públicos primero, después los métodos protegidos y finalmente los métodos privados.

### **Convenciones de los nombres**

- Usa el formato de camelCase, sin subrayados para variables, funciones y nombres de métodos y sus argumentos.
- Usa subrayados para los nombres de opciones y nombres de parámetros.
- Usa namespaces para todas clases.
- Añade el sufijo Interface a los interfaces.
- Utiliza caracteres alfanuméricos y subrayados para nombres de ficheros.

### **Documentación**

- Añadir bloques PHPDoc a todas clases métodos y funciones.

## **6.1.3 PATRÓN DE PROGRAMACIÓN MVC**

Para integrar las funciones estadísticas se desarrolló el módulo de análisis de datos utilizando el framework Symfony a continuación un breve resumen de este patrón.

El término MVC proviene de tres palabras que hoy en día se utilizan mucho dentro del ambiente de desarrollo de software: Model – View – Controller, lo que sería en castellano Modelado, Vista y Controlador. Esta arquitectura permite dividir nuestras aplicaciones en tres grandes capas:

- **Vista:** Todo lo que se refiera a la visualización de la información, el diseño, colores, estilos y la estructura visual en sí de nuestras páginas.
- **Modelado:** Es el responsable de la conexión a la base de datos y la manipulación de los datos mismos. Esta capa está pensada para trabajar con los datos como así también obtenerlos, pero no mostrarlos, ya que la capa de presentación de datos es la vista.
- **Controlador:** Su responsabilidad es procesar y mostrar los datos obtenidos por el Modelado. Es decir, este último trabaja de intermediario entre los otros dos, encargándose también de la lógica de negocio.

Veamos una imagen para tratar de entenderlo mejor:

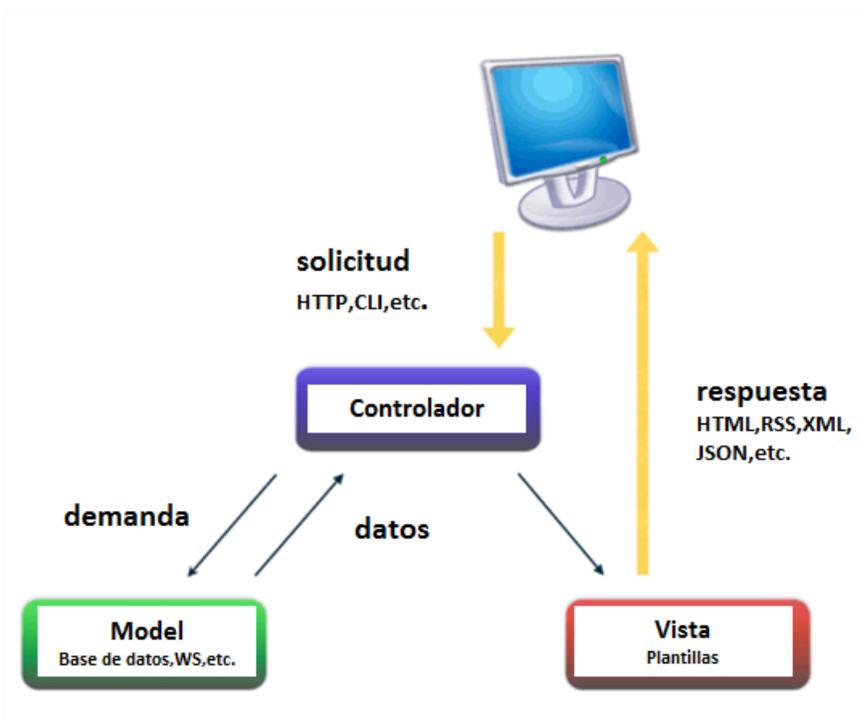


Fig. 6.1 – Diagrama representativo de la Arquitectura MVC en la que se basa Symfony.

El cliente envía una solicitud, ésta es interceptada por el Controlador quien realiza las validaciones necesarias, procesamiento de dichos datos y lógica de negocio asociadas a esa solicitud o petición del cliente. El Controlador envía datos al Modelo, por ejemplo para ser guardados en una base de datos y/o los obtiene dependiendo de la solicitud del usuario para finalmente enviarlos a la Vista a fin de ser mostrados nuevamente al cliente a través de una respuesta.

## 6.2 FUNCIONES ESTADÍSTICAS DEL MÓDULO DE ANÁLISIS DE DATOS

Las funciones contenidas dentro del módulo de datos son:

CÓDIGO	FUNCION ESTADÍSTICA
FE0001	Matriz de Correlación
FE0002	Dendograma
FE0003	Resumen Estadístico
FE0004	Circulo de Correlaciones
FE0005	Centros de Gravedad
FE0006	Intervalos de Confianza
FE0007	Mapa de Correlaciones
FE0008	Análisis de Conglomerados (Clusterización)
FE0009	Diagrama de Caja y Bigote
FE0010	Datos Atípicos
FE0011	Regresión Lineal

Tabla 6.3. Listado de Funciones y su código de identificación

De manera general todos los scripts para las funciones estadísticas funcionan de acuerdo al siguiente procedimiento:

1. Establecen una ruta para el envío de datos al servidor.
2. Obtienen los parámetros necesarios para que la función sea ejecutada correctamente.
3. Realiza la conexión a la base de datos.
4. Ejecuta el script R correspondiente.
5. Retorna el resultado correspondiente en la vista Resultado.html.twig

A continuación se presentan los scripts en código PL/R de cada una de las funciones estadísticas presentes en el módulo de Análisis de Datos.

## 6.2.1 FUNCIÓN MATRIZ DE CORRELACIONES

Código	FE0001
Fecha de Creación	01-09-2015
Objetivo	Calcular la matriz de correlación de un conjunto de datos numéricos.
Nombre de la Función	_plr_matrizcorr
Estructura de la Función	_plr_matrizcorr(arg1)
Tipo de Datos de los Argumentos	arg1: Texto
Descripción de los Argumentos	arg1: Nombre de la tabla que contiene los datos,
Origen de Datos	Tabla temporal que contiene los datos filtrados de un indicador específico
Ejemplo de ejecución de la función	select _plr_matrizcorr('tmp_ind_numero_de_mortalidad_fetal')
Explicación de funcionalidad	La presente función obtiene la matriz de correlación de una tabla de datos, lo que permite detectar cuales variables son dependientes entre sí. Una variable puede tener correlación positiva o negativa con respecto a otra variable, lo que permite determinar su comportamiento. Los valores de la matriz de correlación varían entre 1 y -1, si el valor está cerca de -1, existe correlación negativa, si está cerca de 1, existe correlación positiva, si es cercano a 0, no existe correlación. La función utiliza solo variables numéricas, si en la tabla hay variables categóricas las ignora.
Observaciones o Comentarios: - Poder parametrizar la ubicación de los archivos de imagen que se generan como resultado.	
Script de la Función	
<pre>##Código: FE0001 ##Fecha de Creación: 01-09-15 ##Autores: Willian García, Giovanni Guerrero ##Objetivo: Calcular la matriz de correlación de un conjunto de datos ##numéricos. ##Parámetros y su descripción: arg1: Nombre de la tabla que contiene los ##datos ##Forma de presentación: Reporte que contiene la matriz de correlación de ##las variables en estudio. ##Observaciones o Comentarios: La presente función solo utiliza variables ##numéricas, por lo que si hay variables categóricas en el conjunto de datos ##las ignora.</pre>	

```

##-----INICIO DE SCRIPT-----
CREATE OR REPLACE FUNCTION _plr_matrizcorr(text) RETURNS text AS
$BODY$

##Consulta para obtener datos de indicador de salud.
  select = 'select '
  todos='*'
  from = ' from '
  tabla = arg1

  seleccion = paste(select, todos, from, tabla, sep="");

## Ejecutamos la consulta y la almacenamos en el objeto "datosIndicador";
  datosIndicador = pg.spi.exec (seleccion);
## Modificamos el objeto "datosIndicador" para que seleccione solo las variables numericas.
  datosIndicador= datosIndicador[sapply(datosIndicador,is.numeric)];

## Ejecutamos las respectivas funciones R para obtener la matriz
  matrizcorr<-cor(datosIndicador);

## Creamos la salida de la función, se retornara un reporte.
  reportoutput <- sprintf(" %s ", paste(capture.output(matrizcorr), sep = "", collapse="r"));
  return(reportoutput)

$BODY$
LANGUAGE plr VOLATILE
COST 100;
ALTER FUNCTION _plr_matrizcorr(text)
  OWNER TO postgres;
##-----FIN DE SCRIPT-----

```

Resultado de la función:

	id_establecimiento	edad	muertes_irc
id_establecimiento	1.000000	0.12568299	-0.10744896
edad	0.125683	1.00000000	-0.01276076
muertes_irc	-0.107449	-0.01276076	1.00000000

Tabla 6.4. Función Matriz de Correlaciones.

## 6.2.2 FUNCIÓN DENDOGRAMA

Código	FE0002
Fecha de Creación	01-09-2015
Objetivo	Graficar el dendograma de un conjunto de datos para identificar clústeres, así también obtener los clústeres en un archivo.
Nombre de la Función	_plr_dendograma
Estructura de la Función	_plr_dendograma(arg1,arg2)
Tipo de Datos de los Argumentos	arg1: Texto, arg2: Texto
Descripción de los Argumentos	arg1: Nombre de la tabla que contiene los datos, arg2: Nombre del método a aplicar (puede ser: 'complete', 'single', 'average', 'ward')
Origen de Datos	Tabla temporal que contiene los datos filtrados de un indicador específico
Ejemplo de ejecución de la función	select _plr_dendograma('tmp_ind_numero_de_mortalidad_fetal','complete')
Explicación de funcionalidad	Se obtiene el dendograma de un conjunto de datos, aplicando la librería hclust de R. Así mismo se obtiene un archivo .csv donde se clasifican los clústeres encontrados.
Observaciones o Comentarios: Poder parametrizar la ubicación de los archivos de imagen que se generan como resultado.	
Script de la Función	
<pre> ##Código: FE0002 ##Fecha de Creación: 01-09-15 ##Autores: Willian García, Giovanni Guerrero ##Objetivo: Graficar el dendograma de un conjunto de datos para identificar ##clusters, así también obtener los clusters en un archivo. ##Parámetros y su descripción: arg1: Nombre de la tabla que contiene los ##datos, arg2: Nombre del método a aplicar (puede ser: 'complete' , 'single, ##'ward' y 'average'). ##Forma de presentación: Grafico de dendograma.  ##-----INICIO DE SCRIPT----- <b>CREATE OR REPLACE FUNCTION</b> _plr_dendograma(     text,     text,     integer) <b>RETURNS</b> text <b>AS</b> <b>\$BODY\$</b>  ##Consulta para obtener datos de indicador de salud.</pre>	

```

select = 'select '
todos='*'
tabla = arg1
from = ' from '
seleccion = paste(select, todos, from, tabla, sep="");

## Ejecutamos la consulta y la almacenamos en el objeto "datosIndicador";
datosIndicador = pg.spi.exec (seleccion);

## Lo que obtengamos a partir de aquí se imprimirá en un un archivo .png
dir='/var/www/siigant/src/MINSAL/AnalisisDeDatosBundle/Resources/public/images/graficosr/';
nam= arg1;
ext='.png';
fechahora=format(Sys.time(), "%d.%m.%Y_%H:%M:%S");
newfile=paste(dir,nam,'_',fechahora,ext,sep="");
png(newfile);

## Ejecutamos las respectivas funciones R para obtener el dendograma
dendograma= hclust(dist(datosIndicador), method = arg2);
plot(dendograma);
## la siguiente instrucción separa los clústeres usando 3
rect.hclust(dendograma, k = arg3, border = "red");

## Hasta aquí
dev.off();

## Permisos para archivo
permiso1='chmod go+r';
sys2=paste(permiso1,newfile,sep=' ');
system(sys2);
## cutree corta el el árbol con k clústeres
Grupo <- cutree(dendograma, k = arg3);
NDatos <- cbind(datosIndicador, Grupo);
## Establezco el directorio en donde se quiere grabar el archivo
setwd("/var/www/siigant/src/MINSAL/AnalisisDeDatosBundle/Resources/public/images/graficosr/");
## Se graba el archivo.
archivocsv=paste(arg1,".csv",sep="");
write.csv(NDatos,archivocsv);
## Asignacion de permisos a archivos.
permiso2='chmod 777'
sys3=paste(permiso2,archivocsv,sep=' ');
system(sys3);

## La funcion retorna la ubicacion del grafico y archivo de clusters
respuesta=paste(arg1,'_',fechahora,sep=")
print (respuesta);
$BODY$
LANGUAGE plr VOLATILE
COST 100;
ALTER FUNCTION _plr_dendograma(text, text, integer)
OWNER TO admin;
##-----FIN DE SCRIPT-----

```

Resultado de la función:

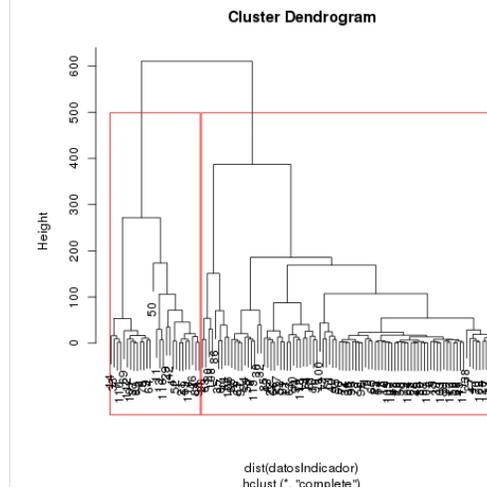


Tabla 6.5. Función Dendrograma.

## 6.2.3 FUNCIÓN RESUMEN ESTADISTICO

Código	FE0003
Fecha de Creación	01-09-2015
Objetivo	Obtener un resumen del cálculo de medidas de posición, tendencia central y dispersión.
Nombre de la Función	_plr_resumenestadistico
Estructura de la Función	_plr_resumenestadistico(arg1,arg2)
Tipo de Datos de los Argumentos	arg1: Texto, arg2: Entero
Descripción de los Argumentos	arg1: Nombre de la tabla que contiene los datos, arg2: Numero de columna de la tabla de datos.
Origen de Datos	Tabla temporal que contiene los datos filtrados de un indicador específico
Ejemplo de ejecución de la función	select _plr_resumenestadistico('tmp_ind_numero_d e_mortalidad_fetal',10)
Explicación de funcionalidad	La función calcula la media, mediana, desviación estándar, desviación típica, deciles, cuartiles y percentiles de un conjunto de datos numéricos.
Observaciones o Comentarios: - Poder parametrizar la ubicación de los archivos de imagen que se generan como resultado.	
Script de la Función	
<pre> ##Código: FE0003 ##Fecha de Creación: 01-09-15 ##Autores: Willian García, Giovanni Guerrero ##Objetivo: Obtener un resumen del cálculo de medidas de posición, tendencia central y dispersión. ##Parámetros y su descripción: arg1: Nombre de la tabla que contiene los datos, arg2: Numero de columna de la tabla de datos. ##Forma de presentación: Reporte con detalle de los cálculos obtenidos.  ##-----INICIO DE SCRIPT----- <b>CREATE OR REPLACE FUNCTION</b> _plr_resumenestadistico(text,integer) <b>RETURNS</b> text <b>AS</b> <b>\$BODY\$</b>  ##Consulta para obtener datos de indicador de salud. <b>select =</b> 'select ' todos='*' tabla = arg1 <b>from =</b> ' from ' </pre>	

```

seleccion = paste(select, todos, from, tabla, sep="");

## Ejecutamos la consulta y la almacenamos en el objeto "datosIndicador";
datosIndicador = pg.spi.exec (seleccion);

## Ejecutamos las respectivas funciones R para obtener el resumen estadístico.
media <- mean(datosIndicador[, arg2]);
mediana <- median(datosIndicador[, arg2]);
desviacion <- sd(datosIndicador[, arg2]);
varianza <- var(datosIndicador[, arg2]);
decil <- quantile(datosIndicador[, arg2],seq(0,1,by=.1));
cuartil <- quantile(datosIndicador[, arg2],seq(0,1,by=.25));
percentil <- quantile(datosIndicador[, arg2],seq(0,1,by=.01));

## Creamos la salida de la función, se retorna un texto con los valores de cada cálculo.
reportoutput <- sprintf(" Media: %s \r\r Mediana: %s \r\r
Desviacion: %s \r\r Varianza: %s \r\r Decil: \r %s \r\r Cuartil: %s
\r\r Percentil: %s", media,
paste(capture.output(median), sep = "", collapse="\r"),
paste(capture.output(sd), sep = "", collapse="\r"),
paste(capture.output(var), sep = "", collapse="\r"),
paste(capture.output(quantile), sep = "", collapse="\r"),
paste(capture.output(quantile), sep = "",
collapse="\r"),paste(capture.output(quantile), sep = "",
collapse="\r")
)
return(reportoutput)

$BODY$
LANGUAGE plr VOLATILE
COST 100;
ALTER FUNCTION _plr_resumenestadistico(text, integer)
OWNER TO postgres;
##-----FIN DE SCRIPT-----

```

Resultado de la función:

```

Media: 68.4416184971098

Mediana: [1] 72

Desviacion: [1] 18.41055

Varianza: [1] 338.9483

Decil:
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
0 43 54 62 67 72 76 80 84 89 102

Cuartil:
0% 25% 50% 75% 100%
0 58 72 82 102

Percentil:
0% 1% 2% 3% 4% 5% 6% 7% 8% 9% 10%
0.00 14.64 20.00 25.92 31.56 33.20 36.00 38.00 39.12 41.76 43.00
11% 12% 13% 14% 15% 16% 17% 18% 19% 20% 21%
44.00 45.00 47.00 48.00 49.00 50.00 51.00 52.00 53.00 54.00 54.00
22% 23% 24% 25% 26% 27% 28% 29% 30% 31% 32%
55.00 56.00 57.00 58.00 58.00 59.28 60.00 61.00 62.00 62.00 62.00

```

Tabla 6.6. Función Resumen Estadístico.

## 6.2.4 FUNCIÓN CÍRCULO DE CORRELACION

Código	FE0004
Fecha de Creación	01-09-2015
Objetivo	Graficar el círculo de correlaciones de un indicador de salud.
Nombre de la Función	_plr_circulocorr
Estructura de la Función	_plr_circulocorr(arg1)
Tipo de Datos de los Argumentos	arg1: Texto
Descripción de los Argumentos	arg1: Nombre de la tabla que contiene los datos
Origen de Datos	Tabla temporal que contiene los datos filtrados de un indicador específico
Ejemplo de ejecución de la función	select _plr_circulocorr('tmp_ind_numero_de_mortalidad_fetal')
Explicación de funcionalidad	La función grafica el circulo de correlaciones de variables numéricas con la finalidad de poder interpretar la correlación entre variables.
Observaciones o Comentarios:	
<ul style="list-style-type: none"> <li>- Poder parametrizar la ubicación de los archivos de imagen que se generan como resultado.</li> </ul>	
Script de la Función	
<pre> ##Código: FE0004 ##Fecha de Creación: 01-09-15 ##Autores: Willian García, Giovanni Guerrero ##Objetivo: Graficar el circulo de correlaciones de un indicador de salud. ##Parámetros y su descripción: arg1: Nombre de la tabla que contiene los ##datos ##Forma de presentación: Grafico del círculo de correlaciones.  ##-----INICIO DE SCRIPT----- <b>CREATE OR REPLACE FUNCTION</b> _plr_circulocorr(     text,     <b>integer</b>) <b>RETURNS</b> text <b>AS</b> <b>\$BODY\$</b>  ##Consulta para obtener datos de indicador de salud. <b>select</b> = 'select ' todos='*' <b>from</b> = ' from ' tabla = arg1 </pre>	

```

seleccion = paste(select, todos, from, tabla, sep=");

## Ejecutamos la consulta y la almacenamos en el objeto "datosIndicador";
datosIndicador = pg.spi.exec (seleccion);

## Lo que obtengamos a partir de aquí se imprimirá en un archivo png
dir='/var/www/siigant/src/MINSAL/AnálisisDeDatosBundle/Resources/public/images/graficosr/';
nam= arg1;
ext='.png';
fechahora=format(Sys.time(), "%d.%m.%Y_%H:%M:%S");
newfile=paste(dir,nam,'_',fechahora,ext,sep=");
png(newfile);

## Ejecutamos las respectivas funciones R para obtener el círculo de
correlaciones
library(FactoMineR)
datosIndicador<-(datosIndicador[sapply(datosIndicador,is.numeric)])
result <- PCA(datosIndicador, scale.unit = TRUE, ncp = arg2, graph = FALSE)
plot(result, axes = c(1, 2), choix = "var", col.var = "blue", new.plot = FALSE)

## Hasta aquí
dev.off();

## Asignación de permisos a archivo
sys1='chmod go+r';
sys2=paste(sys1,"",newfile);
system(sys2);

## La función devuelve el nombre de la tabla, mismo nombre del gráfico
generado en la ubicación:
## /var/www/siigant/src/MINSAL/AnálisisDeDatosBundle/Resources/public/images/graficosr/
respuesta=paste(arg1,'_',fechahora,sep=")
print (respuesta);
$BODY$
LANGUAGE plr VOLATILE
COST 100;
ALTER FUNCTION _plr_circulocorr(text, integer)
##-----FIN DE SCRIPT-----

```

Resultado de la función:

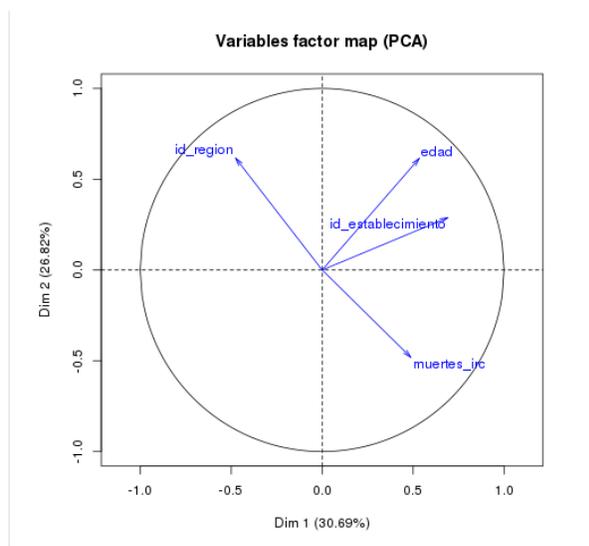


Tabla 6.7. Función Círculo de Correlación.

## 6.2.5 FUNCIÓN CENTROS DE GRAVEDAD

Código	FE0005
Fecha de Creación	01-09-2015
Objetivo	Obtiene los centros de gravedad de los clústeres que se encuentran en un conjunto de datos.
Nombre de la Funcion	_plr_centrosg
Estructura de la Función	_plr_centrosg(arg1,arg2,arg3)
Tipo de Datos de los Argumentos	arg1: Texto, arg2: Texto, Arg3: Entero
Descripción de los Argumentos	arg1: Nombre de la tabla que contiene los datos arg2: Método a aplicar ('complete','average','single','ward') arg3: Tipo de resultado (0:Calcula los centros,1: Grafica histograma de los centros)
Origen de Datos	Tabla temporal que contiene los datos filtrados de un indicador específico
Ejemplo de ejecución de la funcion	select _plr_centrosg('tmp_ind_numero_de_mortalidad_fetal','ward',1)
Explicación de funcionalidad	La funcion calcula los centros de gravedad de los clústeres encontrados en una tabla de datos, así mismo puede obtener el histograma de los centros de gravedad agrupados según el clúster al que pertenecen, esto para facilitar el análisis de los clúster.
Observaciones o Comentarios:	
<ul style="list-style-type: none"> <li>- Poder parametrizar la ubicación de los archivos de imagen que se generan como resultado.</li> </ul>	
Script de la Función	
<pre>##Código: FE0005 ##Fecha de Creación: 01-09-2015 ##Autores: Willian García, Giovanni Guerrero ##Objetivo: Obtiene los centros de gravedad de los clústeres que se ##encuentran en un ##conjunto de datos. ##Parámetros y su descripción: arg1: Nombre de la tabla que contiene los ##datos, arg2: el ##método a aplicar ('complete','average','single','ward'), ##arg3: Resultado que se requiere (0: Calculo de centros de gravedad, 1: ##Histograma de centros de gravedad) ##Forma de presentación: Reporte para el detalle de los cálculos de los ##centros de gravedad o el histograma representativo de los centros de ##gravedad.</pre>	

```

##-----INICIO DE SCRIPT-----
CREATE OR REPLACE FUNCTION _plr_centrosg(
  text,
  text,
  integer)
RETURNS text AS
$BODY$

##Consulta para obtener datos de indicador de salud.
select = 'select '
todos='*'
tabla = arg1
from = ' from '
selection = paste(select, todos, from, tabla, sep="");

## Ejecutamos la consulta y la almacenamos en el objeto "sql";
datosIndicador = pg.spi.exec (selection);
library(rattle)
if (arg3==0){
## Ejecutamos las respectivas funciones R para obtener el dendograma
modelo = hclust(dist(datosIndicador[sapply(datosIndicador,
is.numeric)]), method = arg2)
centros <- centers.hclust(datosIndicador[sapply(datosIndicador,
is.numeric)], modelo, nclust = 3, use.median = FALSE)
## Creamos la salida de la funcion, se retornara un reporte.
reportoutput <- sprintf(" %s ", paste(capture.output(centros), sep
="", collapse="\r"));
return(reportoutput)
}
else {
if (arg3==1){
## Lo que obtengamos a partir de aquí se guardara en un png
dir=/var/www/siigant/src/MINSAL/AnalisisDeDatosBundle/Resources/public/images/graficosr/;
nam= arg1;
ext='.png';
fechahora=format(Sys.time(), "%d.%m.%Y_%H:%M:%S");
newfile=paste(dir,nam,'_',fechahora,ext, sep="");
png(newfile);
modelo = hclust(dist(datosIndicador[sapply(datosIndicador,
is.numeric)]), method = arg2)
centros <- centers.hclust(datosIndicador[sapply(datosIndicador,
is.numeric)], modelo, nclust = 3, use.median = FALSE)
rownames(centros) <- c("Cluster 1", "Cluster 2", "Cluster 3")
barplot(t(centros), beside = TRUE, legend =
colnames(datosIndicador[sapply(datosIndicador, is.numeric)]), main =
"Histograma de Interpretación de Clases",
col = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12), ylim = c(0, 100))
## Hasta aquí
dev.off();
## asignar privilegios de lectura a todos los usuarios de linux
sys1='chmod go+r';
sys2=paste(sys1, " ", newfile);
system(sys2);
respuesta=paste(arg1,'_',fechahora,sep=")
print(respuesta);
}

```

```

else {
print("Error en valor tercer argumento de la funcion");
}
}
}
$BODY$
LANGUAGE plr VOLATILE
COST 100;
ALTER FUNCTION _plr_centrosg(text, text, integer)
OWNER TO postgres;
##-----FIN DE SCRIPT-----

```

Resultado de la función:

	id_region	id_mes	edad	defunciones_cancer_cuello_uterino
[1,]	4.387097	6.290323	62.19355	1.032258
[2,]	3.111111	8.000000	37.55556	1.000000
[3,]	1.000000	8.000000	18.00000	1.000000

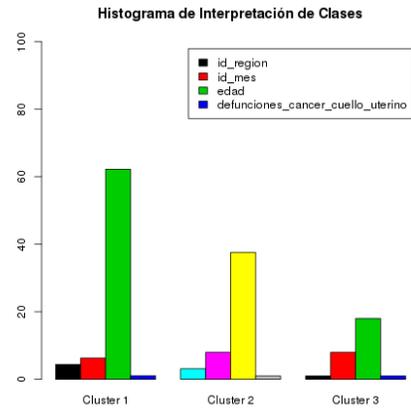


Tabla 6.8. Función Centros de Gravedad.

## 6.2.6 FUNCIÓN INTERVALOS DE CONFIANZA

Código	FE0006
Fecha de Creación	01-09-2015
Objetivo	Encuentra un intervalo de confianza de un conjunto de datos que cumplan con la distribución normal.
Nombre de la Funcion	_plr_intervaloconf
Estructura de la Función	_plr_intervaloconf(arg1,arg2,arg3)
Tipo de Datos de los Argumentos	arg1: Texto, arg2: Texto, Arg3: Double
Descripción de los Argumentos	arg1: Nombre de la tabla que contiene los datos arg2: Numero de columna de la tabla de datos arg3: Nivel de confianza para el intervalo
Origen de Datos	Tabla temporal que contiene los datos filtrados de un indicador específico
Ejemplo de ejecución de la función	select _plr_intervaloconf('tmp_ind_numero_de_mortalidad_fetal',10,0.96)
Explicación de funcionalidad	Se obtiene el cálculo de un intervalo de confianza, de acuerdo al nivel de confianza proporcionado.
Observaciones o Comentarios: - Poder parametrizar la ubicación de los archivos de imagen que se generan como resultado.	
Script de la Función	
<pre> ##Código: FE0006 ##Fecha de Creación: 01-09-2015 ##Autores: Willian García, Giovanni Guerrero ##Objetivo: Encuentra un intervalo de confianza de un conjunto de datos que cumplan con ##la distribución normal. ##Parámetros y su descripción: arg1: Nombre de la tabla que contiene los datos, arg2: ##Numero de columna de la tabla de datos, arg3: Nivel de confianza para el intervalo ##Forma de presentación: Reporte con el detalle de los cálculos obtenidos  ##-----INICIO DE SCRIPT----- <b>CREATE OR REPLACE FUNCTION</b> _plr_intervaloconf(     text,     text,     <b>double precision</b>) <b>RETURNS</b> text <b>AS</b> <b>\$BODY\$</b> </pre>	

```

##Consulta para obtener datos de indicador de salud.
select = 'select '
todos='*'
tabla = arg1
from = ' from '
seleccion = paste(select, todos, from, tabla, sep="");

## Ejecutamos la consulta y la almacenamos en el objeto "datosIndicador";
datosIndicador = pg.sqi.exec (seleccion);
datosIndicador= datosIndicador[, arg2];

## Ejecutamos las respectivas funciones R para encontrar un intervalo
de confianza
intervaloConf<- t.test(datosIndicador,conf.level=arg3);

## Creamos la salida de la funcion, se retornara un reporte.
reportoutput <- sprintf(" Intervalo de Confianza: %s
\\r\\r",paste(capture.output(intervaloConf), sep = "", collapse="\\r")
)
return(reportoutput)

$BODY$
LANGUAGE plr VOLATILE
COST 100;
ALTER FUNCTION _plr_intervaloconf(text, text, double precision)
OWNER TO postgres;
##-----FIN DE SCRIPT-----

```

Resultado de la función:

```

Prueba de Hipotesis:
One Sample t-test

data: datosIndicador
t = -0.51802, df = 61, p-value = 0.6063
alternative hypothesis: true mean is not equal to 56
95 percent confidence interval:
 51.37504 58.72173
sample estimates:
mean of x
 55.04839

```

Tabla 6.9. Función Intervalos de Confianza.

## 6.2.7 FUNCIÓN MAPA DE CORRELACIÓN

Código	FE0007
Fecha de Creación	09-09-2015
Objetivo	Presentar el mapa de correlación de un conjunto de datos numéricos
Nombre de la Función	_plr_mapacorr
Estructura de la Función	_plr_mapacorr(arg1)
Tipo de Datos de los Argumentos	arg1: Texto
Descripción de los Argumentos	arg1: Nombre de la tabla que contiene los datos
Origen de Datos	Tabla temporal que contiene los datos filtrados de un indicador específico
Ejemplo de ejecución de la función	select _plr_mapcorr('tmp_ind_numero_de_mortalidad_fetal')
Explicación de funcionalidad	La presente función genera el mapa de correlación de una tabla de datos, lo que permite detectar cuales variables son dependientes entre sí de manera gráfica. Una variable puede tener correlación positiva o negativa con respecto a otra variable, lo que permite determinar su comportamiento. Los valores de la matriz de correlación varían entre 1 y -1, si el valor está cerca de -1, existe correlación negativa, si está cerca de 1, existe correlación positiva, si es cercano a 0, no existe correlación. La función utiliza solo variables numéricas, si en la tabla hay variables categóricas las ignora.
Observaciones o Comentarios: <ul style="list-style-type: none"> <li>- Poder parametrizar la ubicación de los archivos de imagen que se generan como resultado.</li> </ul>	

## Script de la Función

```
##Codigo: FE0007
##Fecha de Creación: 09-09-15
##Autores: Willian García, Giovanni Guerrero
##Objetivo: Presentar el mapa de correlación de un conjunto de datos
##numéricos
##Estructura: _plr_mapacorr(arg1)
##Parametros y su descripción: arg1: Nombre de la tabla que contiene los
##datos
##Forma de presentación: Grafico

##-----INICIO DE SCRIPT-----
CREATE OR REPLACE FUNCTION _plr_mapacorr(
  text,
  text,
  text,
  text,
  text,
  text,
  text)
RETURNS text AS
$BODY$

##Consulta para obtener datos de indicador de salud.
  select = 'select '
  todos='*'
  from = ' from '
  tabla = arg1

  seleccion = paste(select, todos, from, tabla, sep="");

## Ejecutamos la consulta y la almacenamos en el objeto "datosIndicador";
  datosIndicador = pg.spi.exec (seleccion);

## Modificamos el objeto "datosIndicador" para que seleccione solo las
variables numericas.
  datosIndicador <- datosIndicador[sapply(datosIndicador,is.numeric)];

## Lo que obtengamos a partir de aquí se imprimirá en un un archivo .png
  dir='/var/www/siigant/src/MINSAL/AnalisisDeDatosBundle/Resources/public/images/graficosr/';
  nam= arg1;
  ext='.png';
  fechahora=format(Sys.time(), "%d.%m.%Y_%H:%M:%S");
  newfile=paste(dir,nam,'_',fechahora,ext,sep="");
  png(newfile);

## Ejecutamos las respectivas funciones R para obtener el mapa de correlacion
  data(datosIndicador)
  library(reshape2);
  library(ggplot2);
  graficoqplot<-qplot(x=Var1, y=Var2, data=melt(cor(datosIndicador,
method=arg2)), fill=value, geom="tile", main=arg3, xlab=arg4,
ylab=arg5) +
  scale_fill_gradient2(limits=c(-1, 1)) + theme(legend.position=arg6);
  print(graficoqplot);
## Hasta aquí
```

```

dev.off());

## Asignacion de permisos a archivo
sys1='chmod 777';
sys2=paste(sys1,' ',newfile);
system(sys2);

## La funcion devuelve la ubicacion fisica del mapa generado
respuesta=paste(arg1,'_',fecha hora,sep='')
print (respuesta);

$BODY$
LANGUAGE plr VOLATILE
COST 100;
ALTER FUNCTION _plr_mapacorr(text, text, text, text, text, text)
OWNER TO postgres;
##-----FIN DE SCRIPT-----

```

Resultado de la

función:

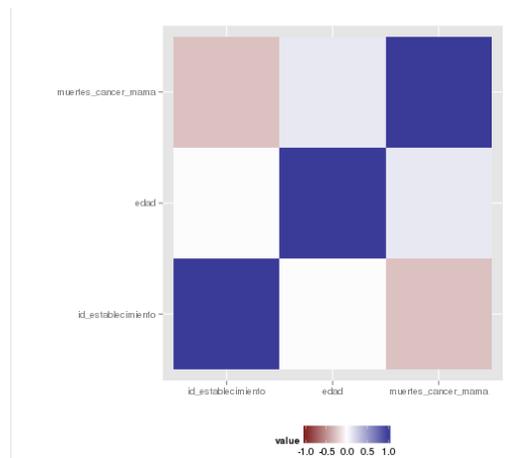


Tabla 6.10. Función Mapa de Correlación.

## 6.2.8 FUNCIÓN ANÁLISIS DE CONGLOMERADOS

Código	FE0008
Fecha de Creación	10-09-2015
Objetivo	Presentar de forma gráfica el análisis de conglomerados (clústeres en el plano)
Nombre de la Función	_plr_clusterizacion
Estructura de la Función	_plr_clusterizacion(arg1)
Tipo de Datos de los Argumentos	arg1: Texto
Descripción de los Argumentos	arg1: Nombre de la tabla que contiene los datos
Origen de Datos	Tabla temporal que contiene los datos filtrados de un indicador específico
Ejemplo de ejecución de la función	select _plr_clusterizacion('tmp_ind_numero_de_mortalidad_fetal')
Explicación de funcionalidad	La presente función busca mediante un gráfico poder representar una realidad que no conseguimos visualizar, una realidad cuya representación original es multidimensional y es imposible que la podamos ver en estado puro. Esto lo hace a través de una agrupación de datos con características similares dentro de un plano.
Observaciones o Comentarios: - Poder parametrizar la ubicación de los archivos de imagen que se generan como resultado.	
Script de la Función	
<pre>##Codigo: FE0008 ##Fecha de Creación: 10-09-15 ##Autores: Willian García, Giovanni Guerrero ##Objetivo: Presentar de forma gráfica el análisis de conglomerados ##(clústeres en el plano) ##Estructura: _plr_clusterizacion(arg1) ##Parametros y su descripcion: arg1: Nombre de la tabla que contiene los</pre>	

```

##datos
##Forma de presentación: Grafico
##-----INICIO DE SCRIPT-----
CREATE OR REPLACE FUNCTION _plr_clusterizacion(
  text,
  integer,
  text,
  text,
  text,
  integer)
RETURNS text AS
$BODY$

##Consulta para obtener datos de indicador de salud.
select = 'select '
  todos='*'
from = ' from '
  tabla = arg1

  seleccion = paste(select, todos, from, tabla, sep="");

##Ejecutamos la consulta y la almacenamos en el objeto "datosIndicador";
  datosIndicador = pg.spl.exec (seleccion);

## Lo que obtengamos a partir de aquí se imprimirá en un un archivo .png
dir=/var/www/siigant/src/MINSAL/AnalisisDeDatosBundle/Resources/public/images/graficosr/;
nam= arg1;
ext='.png!';
fechahora=format(Sys.time(), "%d.%m.%Y_%H:%M:%S");
newfile=paste(dir,nam,'_',fechahora,ext,sep="");
png(newfile);

## Ejecutamos las respectivas funciones R para obtener el analisis de
conglomerados (clusterizacion)
  datosIndicador <-
datosIndicador[sapply(datosIndicador,is.numeric)]; #Solo datos
númerosicos
  datosIndicador = scale(datosIndicador[-1])
  fit <- kmeans(datosIndicador, arg2)
  library(cluster)
  clusplot(datosIndicador, fit$cluster, color=TRUE, shade=TRUE,
labels=arg6, lines=0, main=arg3,
  , xlab=arg4, ylab= arg5)

## Hasta aquí
  dev.off();

## Asignacion de permisos a archivo
  sys1='chmod 777';
  sys2=paste(sys1,' ',newfile);
  system(sys2);

## La funcion devuelve la ubicacion fisica del mapa generado
  respuesta=paste(arg1,'_',fechahora,sep="");
  print (respuesta);

$BODY$

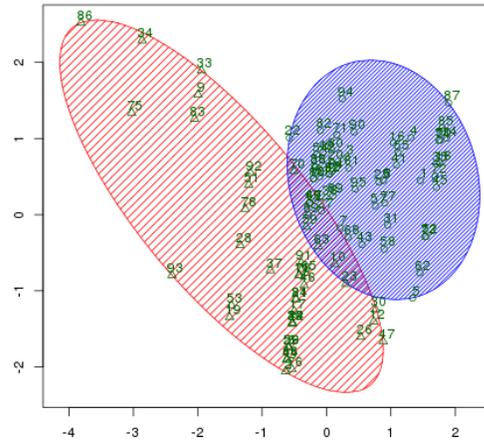
```

```

LANGUAGE plr VOLATILE
COST 100;
ALTER FUNCTION _plr_clusterizacion(text, integer, text, text, text, integer)
OWNER TO postgres;
##-----FIN DE SCRIPT-----

```

Resultado de la función:



These two components explain 77.4 % of the point variability

Tabla 6.11. Función Análisis de Conglomerados.

## 6.2.9 FUNCIÓN DIAGRAMA DE CAJA Y BIGOTE

Código	FE0009
Fecha de Creación	12-09-2015
Objetivo	Presentar de forma gráfica a través del diagrama de caja y bigote varias características importantes de un conjunto de un conjunto de datos al mismo tiempo
Nombre de la Función	_plr_cajabigote
Estructura de la Función	_plr_cajabigote(arg1,arg2,arg3)
Tipo de Datos de los Argumentos	arg1: Texto
Descripción de los Argumentos	arg1: Nombre de la tabla que contiene los datos arg2: Variable que ocupara el eje x arg3: Variable que ocupara el eje y
Origen de Datos	Tabla temporal que contiene los datos filtrados de un indicador específico
Ejemplo de ejecución de la función	select _plr_cajabigote('tmp_ind_numero_de_mortalidad_fetal')
Explicación de funcionalidad	La presente función busca presentar de forma visual varias características importantes de un conjunto de un conjunto de datos al mismo tiempo, tales como la dispersión y simetría.
Observaciones o Comentarios: - Poder parametrizar la ubicación de los archivos de imagen que se generan como resultado.	
Script de la Función	
<b>##Codigo:</b> FE0010 <b>##Fecha de Creación:</b> 12-09-15 <b>##Autores:</b> Willian García, Giovanni Guerrero <b>##Objetivo:</b> Presentar de forma gráfica a través del diagrama de caja y ##bigote varias características	

```

importantes de un conjunto de un conjunto de ##datos al mismo tiempo
##Estructura: _plr_cajabigote(arg1,arg2,arg3)
##Parametros y su descripcion: arg1: Nombre de la tabla que contiene los ##datos, arg2: Variable
que ocupara el eje x, arg3: Variable que ocupara el ##eje y
##Forma de presentación: Grafico
##-----INICIO DE SCRIPT-----
CREATE OR REPLACE FUNCTION _plr_cajabigote(
    text,
    integer,
    text,
    text,
    text)
RETURNS text AS
$BODY$

##Consulta para obtener datos de indicador de salud.
select = 'select '
    todos='*'
from = ' from '
    tabla = arg1

    seleccion = paste(select, todos, from, tabla, sep="");

## Ejecutamos la consulta y la almacenamos en el objeto "datosIndicador";
    datosIndicador = pg.spi.exec (seleccion);

## Lo que obtengamos a partir de aquí se imprimirá en un un archivo .png
    dir='/var/www/siigant/src/MINSAL/AnalisisDeDatosBundle/Resources/public/images/graficosr/';
    nam= arg1;
    ext='.png';
    fechahora=format(Sys.time(), "%d.%m.%Y_%H:%M:%S");
    newfile=paste(dir,nam,'_',fechahora,ext,sep="");
    png(newfile);

## Ejecutamos las respectivas funciones R para obtener el diagrama de
caja y bigote
    datosIndicador <-
datosIndicador[sapply(datosIndicador,is.numeric)]; #Solo datos
numéricos
    if (arg2==0){
        boxplot(as.data.frame(datosIndicador), main=arg3, xlab=arg4,
ylab=arg5, outline=F)
    }
else{
        boxplot(as.data.frame(datosIndicador), main=arg3, xlab=arg4,
ylab=arg5, outline=T)
    }
## Hasta aquí
    dev.off();

## Asignacion de permisos a archivo
    sys1='chmod 777';
    sys2=paste(sys1,' ',newfile);
    system(sys2);

## La funcion devuelve la ubicacion fisica del grafico generado
    respuesta=paste(arg1,'_',fechahora,sep="");

```

```
print (respuesta);
```

```
$BODY$
```

```
LANGUAGE plr VOLATILE
```

```
COST 100;
```

```
ALTER FUNCTION _plr_cajabigote(text, integer, text, text, text)
```

```
OWNER TO admin;
```

```
##-----FIN DE SCRIPT-----
```

Resultado de la función:

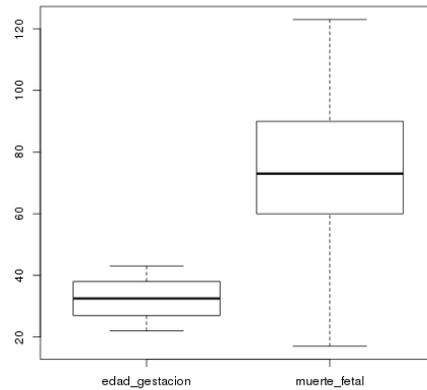


Tabla 6.12. Función Diagrama de Caja y Bigote.

## 6.2.10 FUNCIÓN DATOS ATÍPICOS O MAL REPRESENTADOS

Código	FE0010
Fecha de Creación	13-09-2015
Objetivo	Detectar el conjunto de datos atípicos o mal representados dentro del conjunto de datos que generan un indicador
Nombre de la Función	_plr_atipicos
Estructura de la Función	_plr_atipicos(arg1)
Tipo de Datos de los Argumentos	arg1: Texto
Descripción de los Argumentos	arg1: Nombre de la tabla que contiene los datos
Origen de Datos	Tabla temporal que contiene los datos filtrados de un indicador específico
Ejemplo de ejecución de la función	select _plr_atipicos('tmp_ind_numero_de_mortalidad_fetal')
Explicación de funcionalidad	La presente función busca mediante un gráfico poder representar una realidad que no conseguimos visualizar, una realidad cuya representación original es multidimensional y es imposible que la podamos ver en estado puro. Esto lo hace a través de una agrupación de datos con características similares dentro de un plano.
Observaciones o Comentarios: - Poder parametrizar la ubicación de los archivos de imagen que se generan como resultado.	
Script de la Función	
<pre>##Codigo: FE0010 ##Fecha de Creación: 13-09-15 ##Autores: Willian García, Giovanni Guerrero ##Objetivo: Detectar el conjunto de datos atípicos o mal representados ##dentro del conjunto de datos que generan un indicador ##Estructura: _plr_atipicos(arg1)</pre>	

```

##Parametros y su descripcion: arg1: Nombre de la tabla que contiene los ##datos
##Forma de presentación: Datos
##-----INICIO DE SCRIPT-----
CREATE OR REPLACE FUNCTION _plr_atipicos(text)
RETURNS text AS
$BODY$

##Consulta para obtener datos de indicador de salud.
select = 'select '
  todos='*'
from = ' from '
  tabla = arg1

  seleccion = paste(select, todos, from, tabla, sep="");

## Ejecutamos la consulta y la almacenamos en el objeto "datosIndicador";
  datosIndicador = pg.sql.exec (seleccion);

## Ejecutamos las respectivas funciones R para obtener datis atipicos
  haciendo uso del diagrama de caja y bigote
  datosIndicador=datosIndicador[sapply(datosIndicador,is.numeric)]
  atipicos <- boxplot(datosIndicador,plot=F)
  salida<-c(atipicos$names[atipicos$group],atipicos$out)

## Creamos la salida de la funcion, se retorna un texto con los
  valores de cada calculo.
  print(salida)

$BODY$
LANGUAGE plr VOLATILE
COST 100;
ALTER FUNCTION _plr_atipicos(text)
  OWNER TO postgres;
##-----FIN DE SCRIPT-----

```

Resultado de la función:

	edad	muertes_irc
Datos atipicos	101, 1102	25, 45, 125

Tabla 6.13. Función Datos Atípicos.

## 6.2.11 FUNCIÓN REGRESIÓN LINEAL

Código	FE0011
Fecha de Creación	14-09-2015
Objetivo	Presentar de forma gráfica la regresión lineal presentada por el conjunto de datos que generan un indicador.
Nombre de la Función	_plr_regresionlineal
Estructura de la Función	_plr_regresionlineal(arg1,arg2,arg3,arg4)
Tipo de Datos de los Argumentos	arg1: Texto arg2: Texto arg3: Texto arg4: Texto
Descripción de los Argumentos	arg1: Nombre de la tabla que contiene los datos  arg2: variable del conjunto de datos a considerar dentro del análisis de regresión lineal  arg3: variable del conjunto de datos a considerar dentro del análisis de regresión lineal  arg4: variable del conjunto de datos a considerar dentro del análisis de regresión lineal
Origen de Datos	Tabla temporal que contiene los datos filtrados de un indicador específico
Ejemplo de ejecución de la función	<pre>select _plr_regresion lineal("tmp_ind_numero_de_mortalidad_fetal", "edad")</pre>
Explicación de funcionalidad	La presente función busca mediante un gráfico poder representar el comportamiento lineal del conjunto de datos que forman un indicador de salud y de esta manera poder predecir su comportamiento a futuro.

Observaciones o Comentarios:

- Poder parametrizar la ubicación de los archivos de imagen que se generan como resultado.

Script de la Función

```
##Codigo: FE0011
##Fecha de Creación: 14-09-15
##Autores: Willian García, Giovanni Guerrero
##Objetivo: Presentar de forma gráfica la regresión lineal presentada por el ##conjunto de datos que
generan un indicador.
##Estructura: _plr_regresionlineal(arg1,arg2,arg3,arg4)
##Parametros y su descripcion: arg1: Nombre de la tabla que contiene los ##datos
##Forma de presentación: Grafico
##-----INICIO DE SCRIPT-----
CREATE OR REPLACE FUNCTION _plr_regresionlineal(
  text,
  text,
  text,
  text,
  text,
  text,
  text)
RETURNS text AS
$BODY$

##Consulta para obtener datos de indicador de salud.
  select = 'select '
  todos='*'
  from = ' from '
  tabla = arg1

  seleccion = paste(select, todos, from, tabla, sep="");

## Ejecutamos la consulta y la almacenamos en el objeto "datosIndicador";
  datosIndicador = pg.spi.exec (seleccion);

## Lo que obtengamos a partir de aquí se imprimirá en un un archivo .png
  dir='/var/www/siigant/src/MINSAL/AnálisisDeDatosBundle/Resources/public/images/graficosr/';
  nam= arg1;
  ext='.png';
  fechahora=format(Sys.time(), "%d.%m.%Y_%H:%M:%S")
  newfile=paste(dir,nam,'_',fechahora,ext,sep="");
  png(newfile);

## Ejecutamos las respectivas funciones R para obtener la regresion
  datosIndicador <-
datosIndicador[sapply(datosIndicador,is.numeric)]; #Solo datos
numéricos
  y<-datosIndicador[,c(arg2)]
  x<-datosIndicador[,c(arg3)]
  regresion <- lm(y ~ x,data=datosIndicador)
  plot(datosIndicador[,c(arg3)],datosIndicador[,c(arg2)],
main=arg4,xlab=arg5,ylab=arg6)
  abline(regresion,col=2)
```

```

## Hasta aquí
dev.off();

## Asignacion de permisos a archivo
sys1='chmod 777 ';
sys2=paste(sys1,newfile);
system(sys2);

## La funcion devuelve el nombre de la tabla que corresponde al grafico generado
respuesta=paste(arg1,'_',fechahora,sep='')
print (respuesta);

$BODY$
LANGUAGE plr VOLATILE
COST 100;
ALTER FUNCTION _plr_regresionlineal(text, text, text, text, text, text)
OWNER TO postgres;

##-----FIN DE SCRIPT-----

```

Resultado de la función:

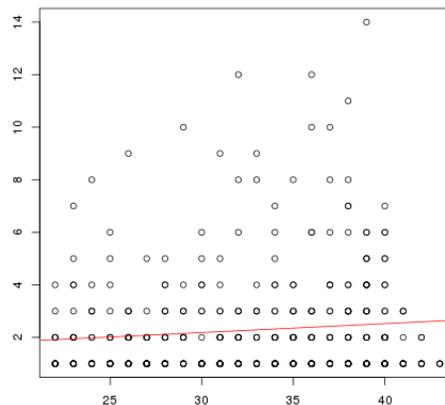


Tabla 6.14. Función Regresión Lineal.

## 6.3 INTEGRACIÓN CON EL SISTEMA DE INDICADORES

### Pasos previos para integrar el módulo de análisis estadístico:

1. Instalar R statistics versión 3.1<sup>27</sup>
2. Instalar la extensión PLR en el gestor de la base de datos del SIIG.<sup>28</sup>
3. Crear las nuevas tablas que utilizara el modulo en la base de datos.
4. Ejecutar los scripts correspondientes a las funciones estadísticas desarrolladas en la base de datos del SIIG.
5. Realizar pruebas de las funciones estadísticas en la base de datos del SIIG.

### Pasos para integrar el modulo al SIIG.

1. Pegar el código del módulo análisis de datos (AnalisisDeDatosBundle) en la siguiente ubicación del SIIG:

```
src\MINSAL\
```

2. Registrar el modulo en el archivo appKernel.php de la siguiente forma:

```
new MINSAL\AnalisisDeDatosBundle\AnalisisDeDatosBundle(),
```

3. Crear la ruta correspondiente del bundle análisis de datos en el archivo Routing.yml

```
Analisis_De_Datos:  
resource:@AnalisisDeDatosBundle/Controller/"  
type: annotation  
prefix: /
```

4. Importar el nombre del archivo admin.xml en el archivo config.yml del SIIG.

```
- { resource: @AnalisisDeDatosBundle/Resources/config/admin.xml }
```

5. Instalar los assets con el siguiente comando.

```
php app/console assets:install web --symlink
```

6. Ingresar al SIIG y probar el modulo análisis de datos.

---

<sup>27</sup> [http://wiki.salud.gob.sv/wiki/Instalación\\_de\\_R\\_y\\_RStudio](http://wiki.salud.gob.sv/wiki/Instalación_de_R_y_RStudio)

<sup>28</sup> [http://wiki.salud.gob.sv/wiki/Instalación\\_de\\_R\\_y\\_RStudio](http://wiki.salud.gob.sv/wiki/Instalación_de_R_y_RStudio)

## **6.4 DOCUMENTACIÓN**

La documentación nos sirve para tener una mejor apreciación del módulo y para la solución de problemas, así como de apoyo para la utilización de la aplicación. El presente módulo presenta la siguiente documentación de soporte:

### **MANUAL DE USUARIO**

Este manual nos proporciona una guía en la utilización de la herramienta de análisis de datos, se detalla cada formulario correspondiente a cada función estadística, filtración de datos de las tablas temporales del almacén, explicación de los parámetros de cada función y la ayuda correspondiente a cada función.

Para ver este manual buscarlo en el CD de soporte del módulo con el nombre Manual de Usuario.

### **MANUAL TÉCNICO**

Con la finalidad de apoyar al mantenimiento de la aplicación o si se desea agregar o modificar características del módulo, así como solucionar posibles problemas se aporta este manual. El contenido del manual incluye el detalle de los orígenes de datos, las funciones estadísticas desarrolladas, las diferentes vistas, el controlador y funciones de jquery para la parte dinámica de la aplicación. También se detalla la ubicación de cada archivo que constituye el código del módulo.

Para ver este manual buscarlo en el CD de soporte del módulo con el nombre Manual de Técnico.

## 6.5 PLAN DE IMPLEMENTACIÓN

Las actividades a considerar para la implementación del módulo de análisis de datos con su respectivo tiempo son las siguientes:

Actividad	Tiempo
Capacitación del personal	5 días
Presentación del proyecto	1 día
Instalación de componentes	1 día
Configuración de componentes	1 día
Carga de scripts y tablas a base de datos	1 día
Pruebas de implementación	5 días
Realización de pruebas	15 días
Análisis de pruebas	5 días

Tabla 6.15. Actividades plan de implementación.

- **Capacitación del personal:** Consiste en la preparación de los usuarios que utilizaran el módulo de análisis de datos.

Actividad	Acción
Generalidades	Beneficios del módulo Explicación general del módulo
Acceso al módulo	Ingreso al módulo
Interfaz del módulo	Verificar interfaz del módulo Explicar opciones de menú
Pruebas de captura de datos	Ingresar datos de prueba
Pruebas de funciones estadísticas	Ejecutar funciones estadísticas Verificar resultados
Otros	Aclaración de dudas.

Tabla 6.16. Actividades capacitación.

- **Presentación del proyecto:** Consiste en realizar una presentación del módulo a los usuarios beneficiarios.
- **Instalación de componentes:** Incluye los componentes necesarios para el correcto funcionamiento del módulo de análisis de datos.

Para esto considerar lo siguiente:

- Instalación de R Statistics en servidor.
- Instalación de extensión PLR.

- **Configuración de componentes:** Se deben realizar las configuraciones necesarias para el nuevo módulo, esto incluye:
  - Agregar el código del módulo al SIIG.
  - Realizar las configuraciones necesarias para integrar el modulo al SIIG.
  
- **Carga de Scripts y tablas a la base de datos:** Se deben cargar las funciones estadísticas desarrolladas en la base de datos, que permiten la conexión con R statistics, así como las nuevas tablas que serán parte del SIIG.
  
- **Pruebas de Implementación:** Las pruebas de implementación se llevaran a cabo para verificar la correcta integración del módulo. Estas pruebas deben ser de acuerdo a los lineamientos del Sistema Integrado de Indicadores Gerenciales.
  
- **Realización de pruebas:** Consiste en probar la funcionalidad del módulo, es decir la correcta obtención de resultados de las funciones estadísticas desarrolladas,
  
- **Análisis de pruebas:** Para finalizar se debe analizar todas las pruebas realizadas para garantizar el correcto funcionamiento del módulo integrado.

**Equipo para la implementación del módulo.**

<b>Equipo</b>	<b>Descripción</b>
10 computadoras	Para capacitar a los usuarios se debe disponer de computadoras para las prácticas respectivas.
1 Proyector	Para la presentación del proyecto, así como para la capacitación de los usuarios.
1 Laptop	Se necesita para impartir las capacitaciones a los usuarios.
Guía practica	Con la finalidad de impartir los temas respectivos del módulo.

Tabla 6.17. Equipo requerido para implementación.

## CONCLUSIONES

Finalizado el proyecto “Desarrollo e Integración de Funciones Estadísticas para el Análisis de Datos en el Sistema Integrado de Indicadores Gerenciales del Ministerio de Salud”, se ha llegado a las siguientes conclusiones:

- ✓ El Impacto que representa para el Ministerio de Salud el uso del análisis de datos será positivo y permitirá realizar una mejora significativa en la toma de decisiones, ya que podrán contar con datos más exactos y precisos que les permitan tener una mejor apreciación de la situación actual y futura, lo que les será útil al momento de definir estrategias y realizar planificaciones.
- ✓ Hay diferentes tipos de funciones estadísticas, las cuales se pueden clasificar de acuerdo a la tarea o tipo de conocimiento que se obtiene de ellas. Es importante prestar mucha atención al tipo de conocimiento que se quiere obtener para así fácilmente poder elegir la función más adecuada que satisfaga las necesidades de las unidades beneficiadas.
- ✓ Al aplicar análisis de datos al Almacén de Datos de Indicadores de Salud, el Ministerio de Salud está haciendo uso de técnicas vanguardistas para apoyar la toma de decisiones y a la vez sustentar de manera sólida su planificación a corto, mediano y largo plazo.

## RECOMENDACIONES

- ✓ Realizar un estudio previo a los datos antes de utilizar las funciones estadísticas, esto con la finalidad que los resultados obtenidos sean coherentes y proporcionen la información correcta.
  
- ✓ Verificar la estructura de los datos en general a analizar, estos no deben presentar campos en blanco y se debe verificar el tipo de dato de cada variable, ya que la mayoría de los métodos estadísticos solo utilizan variables numéricas, con esto se evitan errores en tiempo de ejecución de la función estadística que se aplique.
  
- ✓ Considerar agregar nuevos datos de los indicadores como parte de las fuentes de datos que alimentan a las funciones estadísticas a desarrollar, a modo de permitir nuevos resultados, complementar los que ya existen y de esta manera generar nuevos beneficios para las personas que hagan uso del Módulo de Análisis de Datos como parte del desempeño de sus actividades.
  
- ✓ Evaluar periódicamente los resultados que se obtienen con el uso de las funciones estadísticas, auxiliándose en los resultados que estas brindan para apoyar la toma de decisiones; esto con el fin de buscar nuevas oportunidades de mejora o encontrar alguna carencias considerable dentro de su funcionalidad.

## GLOSARIO

- **Almacenes de Datos (Data Warehouse):** Es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.
- **Análisis Multidimensional:** Es un análisis de los datos de proceso que agrupa los datos en dos o más categorías: dimensiones de los datos y categorías.
- **Archivo CSV:** Son un tipo de documento en formato abierto sencillo para representar datos en forma de tabla, en las que las columnas se separan por comas (o punto y coma en donde la coma es el separador decimal: Argentina, Brasil...) y las filas por saltos de línea. Los campos que contengan una coma, un salto de línea o una comilla doble deben ser encerrados entre comillas dobles.
- **Bases de Datos:** Es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.
- **BID:** El *Banco Interamericano de Desarrollo* (BID) es una organización financiera internacional con sede en la ciudad de Washington D.C. (Estados Unidos), y creada en el año de 1959 con el propósito de financiar proyectos viables de desarrollo económico, social e institucional y promover la integración comercial regional en el área de América Latina y el Caribe.
- **Categorización:** Es el proceso por el cual especificamos cuáles serán las categorías de la variable que interesan. A su vez, las categorías o valores son las diferentes posibilidades de variación que una variable puede tener.
- **CDC:** Los *Centros para el Control y la Prevención de Enfermedades* (CCPEEU) (en inglés Centers for Disease Control and Prevention, CDC) son una agencia del Departamento de Salud y Servicios Humanos de los Estados Unidos cuya responsabilidad a nivel nacional radica en el desarrollo y la aplicación de la prevención y control de enfermedades, salud ambiental y la realización de actividades de educación y promoción de la salud.
- **Código Libre (Open Source):** Es la expresión con la que se conoce al software distribuido y desarrollado libremente.

- **Clustering:** Procedimiento de agrupación de una serie de vectores de acuerdo con un criterio. Esos criterios son por lo general distancia o similitud. La cercanía se define en términos de una determinada función de distancia, como la euclídea, aunque existen otras más robustas o que permiten extenderla a variables discretas. Generalmente, los vectores de un mismo grupo (o clústers) comparten propiedades comunes.
- **Cubo OLAP:** Es una base de datos multidimensional, en la cual el almacenamiento físico de los datos se realiza en un vector multidimensional.
- **DTIC:** Dirección de Tecnologías de Información y Comunicaciones del Ministerio de Salud. Entidad encargada de administrar y regular todo lo relacionado a las Ciencias de la Información dentro del Ministerio de Salud.
- **Estadística Bayesiana:** Es un subconjunto del campo de la estadística en la que la evidencia sobre el verdadero estado del mundo se expresa en términos de grados de creencia o, más específicamente, las probabilidades bayesianas. Tal interpretación es sólo una de una serie de interpretaciones de la probabilidad y hay otras técnicas estadísticas que no se basan en "grados de creencia".
- **Flujo de datos:** Son movimientos de datos en una determinada dirección, desde un origen hasta un destino. Es un paquete de datos.
- **Granularidad:** El concepto de granularidad parte del principio que es más fácil reutilizar unidades más pequeñas dado, que de este modo, es posible seleccionar aquellas partes que nos interesan y descartar aquellas que no son adecuadas en el contexto donde nos encontramos. Además la granularidad describe el nivel de detalle de la base de datos en data warehouse.
- **Indicador de Salud:** Es una medida que se puede usar para ayudar a describir una situación existente o para medir cambios o tendencias a lo largo de un periodo de tiempo, estos indicadores pueden ser cualitativos y cuantitativos y están relacionados al área de la salud humana.
- **Metadatos:** En general, un grupo de metadatos se refiere a un grupo de datos, llamado recurso. El concepto de metadatos es análogo al uso de índices para localizar objetos en vez de datos. Por ejemplo, en una biblioteca se usan fichas que especifican autores, títulos, casas editoriales y lugares para buscar libros. Así, los metadatos ayudan a ubicar datos.

- **Middleware:** Es un software que asiste a una aplicación para interactuar o comunicarse con otras aplicaciones, o paquetes de programas, redes, hardware y/o sistemas operativos. Éste simplifica el trabajo de los programadores en la compleja tarea de generar las conexiones y sincronizaciones que son necesarias en los sistemas distribuidos. De esta forma, se provee una solución que mejora la calidad de servicio, así como la seguridad, el envío de mensajes, la actualización del directorio de servicio, etc.
- **Ministerio de Salud de El Salvador (MINSAL):** Es la instancia del Estado rectora en materia de salud que garantiza a los habitantes de la República de El Salvador la cobertura de servicios oportunos e integrales.
- **Modelo de datos EAV:** Es un modelo de datos que describe entidades donde el número de atributos que las describen puede ser bastante amplio.
- **OLTP:** Es la sigla en inglés de *Procesamiento de Transacciones En Línea* (OnLine Transaction Processing) es un tipo de procesamiento que facilita y administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones (gestor transaccional). Los paquetes de software para OLTP se basan en la arquitectura cliente-servidor ya que suelen ser utilizados por empresas con una red informática distribuida.
- **OPS:** La *Organización Panamericana de la Salud* (OPS) es el organismo especializado de salud del sistema interamericano, encabezado por la Organización de los Estados Americanos (OEA), y también está afiliada a la Organización Mundial de la Salud (OMS), desde 1949, de manera que forma parte igualmente del sistema de las Naciones Unidas.
- **Patrón “Atributos y Tipos” (EAV por sus siglas en inglés):** permite tener un dominio detallado sobre todos los atributos que podamos asignarle a cualquier elemento que es almacenado, en pocas palabras, cada registro habla por si solo y no tendrá libre interpretación, las aplicaciones que hacen uso de estos valores serán dominados directamente por el manejador de base de datos, y al final todo este proceso conlleva a una simple solución de auto-documentación e integridad de los datos.
- **PSPP:** Aplicación de software libre para el análisis de datos. Se presenta en modo gráfico y está escrita en el lenguaje de programación C. Usa la biblioteca científica *GNU* para sus rutinas matemáticas, y *plotutils* para la generación de gráficos. Es un reemplazo libre para el software propietario SPSS. Proporciona funcionalidades básicas como: frecuencias, tablas cruzadas, comparación de media; regresión lineal, fiabilidad,

reordenamiento de datos, pruebas no paramétricas, factor de análisis entre otras características.

- **RIISS:** Acrónimo para referirse a las Redes Integrales e Integradas de Servicios de Salud del Ministerio de Salud de El Salvador.
- **SIIG/eTAB:** El Sistema Integrado de Indicadores Gerenciales, es un proyecto Open Source, que nace a inicios del año 2011 como una iniciativa de evaluar el progreso de la Reforma de Salud, apoyados por la Organización Panamericana de la Salud (OPS).
- **Sistema de Información Gerencial:** Son una colección de sistemas de información que interactúan entre sí y que proporcionan información tanto para las necesidades de las operaciones como de la administración.
- **SPSS:** Programa estadístico informático muy usado en las ciencias sociales y las empresas de investigación de mercado. Originalmente SPSS fue creado como el acrónimo de Statistical Package for the Social Sciences aunque también se ha referido como "Statistical Product and Service Solutions" (Pardo, A., & Ruiz, M.A., 2002, p. 3). Actualmente, la parte SPSS del nombre completo del software (*IBM SPSS*) no es acrónimo de nada. Es uno de los programas estadísticos más conocidos teniendo en cuenta su capacidad para trabajar con grandes bases de datos y un sencillo interface para la mayoría de los análisis. En la versión 12 de SPSS se podían realizar análisis con 2 millones de registros y 250.000 variables. El programa consiste en un módulo base y módulos anexos que se han ido actualizando constantemente con nuevos procedimientos estadísticos. Cada uno de estos módulos se compra por separado

## REFERENCIAS

### BIBLIOGRAFICAS

WHITTEN, J.L., BENTLEY, L.D., BARLOW, V.M. : Análisis y Diseño de Sistemas de Información, 3ª Ed., Irwin, 2003.

SENN, J. A. *Análisis y diseño de sistemas de información*. Segunda Edición, McGraw-Hill, Abril 2000.

### SITIOS WEB

#### **Página de información informática general del Ministerio de Salud de El Salvador**

Título del Artículo Consultado: UnoTAB

Fecha de Publicación: Desconocido

Fecha de Consulta: 28 de marzo de 2015

<http://wiki.salud.gob.sv/wiki/UnoTab>

#### **Arquitectura de un Almacén de Datos.**

Título del Artículo Consultado: Arquitectura Almacén de Datos.

Fecha de Publicación: Desconocido

Fecha de Consulta: 10 de Junio de 2015

<http://anabuigues.com/2010/03/05/arquitectura-de-un-data-warehouse/>

#### **Blog Carbarez'sWeblog**

Título del Artículo Consultado: Caja Negra

Fecha de Publicación: Desconocido

Fecha de Consulta: 3 de Abril de 2015

<http://carolbr.wordpress.com/caja-negra/>

#### **Sitio web Search Data Center en Español**

Título del Artículo Consultado: Análisis de Datos

Fecha de Publicación: 20 de Noviembre de 2012

Fecha de Consulta: 1 de Abril de 2015

<http://searchdatacenter.techtarget.com/es/definicion/Analisis-de-Datos>

**Sitio web OpenCourseWare de la Universidad de Sevilla**

Título del Artículo Consultado: Análisis de Datos en la Investigación Educativa / Bloque I

Subtítulo: El concepto de análisis estadístico de datos

Fecha de Publicación: Desconocido

Fecha de Consulta: 1 de Abril de 2015

[http://ocwus.us.es/metodos-de-investigacion-y-diagnostico-en-educacion/analisis-de-datos-en-la-investigacion-educativa/Bloque\\_I/page\\_03.htm](http://ocwus.us.es/metodos-de-investigacion-y-diagnostico-en-educacion/analisis-de-datos-en-la-investigacion-educativa/Bloque_I/page_03.htm)

**Sitio web OpenCourseWare de la Universidad de Sevilla**

Título del Artículo Consultado: Análisis de Datos en la Investigación Educativa / Bloque I

Subtítulo: Análisis de Datos

Fecha de Publicación: Desconocido

Fecha de Consulta: 1 de Abril de 2015

[http://ocwus.us.es/metodos-de-investigacion-y-diagnostico-en-educacion/analisis-de-datos-en-la-investigacion-educativa/Bloque\\_I/page\\_07.htm](http://ocwus.us.es/metodos-de-investigacion-y-diagnostico-en-educacion/analisis-de-datos-en-la-investigacion-educativa/Bloque_I/page_07.htm)

**Sitio web OpenCourseWare de la Universidad de Sevilla**

Título del Artículo Consultado: Análisis de Datos en la Investigación Educativa / Bloque I

Subtítulo: La Informática en el Análisis Estadístico de Datos

Fecha de Publicación: Desconocido

Fecha de Consulta: 1 de Abril de 2015

[http://ocwus.us.es/metodos-de-investigacion-y-diagnostico-en-educacion/analisis-de-datos-en-la-investigacion-educativa/Bloque\\_I/page\\_10.htm](http://ocwus.us.es/metodos-de-investigacion-y-diagnostico-en-educacion/analisis-de-datos-en-la-investigacion-educativa/Bloque_I/page_10.htm)

**Sitio web DeGerencia.com**

Título del Artículo Consultado: Toma de decisiones

Fecha de Publicación: Desconocido

Fecha de Consulta: 8 de Abril de 2015

[http://www.degerencia.com/tema/toma\\_de\\_decisiones](http://www.degerencia.com/tema/toma_de_decisiones)

**Sitio web Epicentro**

Título del Artículo Consultado: Indicadores en salud

Fecha de Publicación: Desconocido

Fecha de Consulta: 13 de Abril de 2015

<http://escuela.med.puc.cl/recursos/recepidem/insintrod6.htm>

**Sitio web de la Organización Panamericana de la Salud**

Título del Artículo Consultado: Glosario de Indicadores de salud

Fecha de Publicación: Desconocido

Fecha de Consulta: 13 de Abril de 2015

<http://www1.paho.org/Spanish/SHA/CoreData/Tabulator/glossary.htm>

---

---

# ANEXOS

## **SINOPSIS**

Documentos complementarios que permiten ampliar algunos temas relevantes expuestos en el contenido de este trabajo.

---

---

# ANEXO 1

## INVESTIGACIÓN DETALLADA SOBRE FUNCIONES ESTADÍSTICAS

### FUNCIONES ESTADÍSTICAS

#### Medidas de posición<sup>29</sup>

Las medidas de posición nos facilitan información sobre la serie de datos que estamos analizando. La descripción de un conjunto de datos, incluye como un elemento de importancia la ubicación de éstos dentro de un contexto de valores posible. Una vez definidos los conceptos básicos en el estudio de una distribución de frecuencias de una variable, estudiaremos las distintas formas de resumir dichas distribuciones mediante medidas de posición (o de centralización), teniendo presente el error cometido en el resumen mediante las correspondientes medidas de dispersión.

Se trata de encontrar unas medidas que sinteticen las distribuciones de frecuencias. En vez de manejar todos los datos sobre las variables, tarea que puede ser pesada, podemos caracterizar su distribución de frecuencias mediante algunos valores numéricos, eligiendo como resumen de los datos un valor central alrededor del cual se encuentran distribuidos los valores de la variable

Pero estas medidas de posición de una distribución de frecuencias han de cumplir determinadas condiciones para que lean verdaderamente representativas de la variable a la que resumen. Toda síntesis de una distribución se considerara como operativa si intervienen en su determinación todos y cada uno de los valores de la distribución, siendo única para cada distribución de frecuencias y siendo siempre calculable y de fácil obtención. A continuación se describen las medidas de posición más comunes utilizadas en estadística, como lo son:

**Cuartiles:** Hay 3 cuartiles que dividen a una distribución en 4 partes iguales: primero, segundo y tercer cuartil.

**Deciles:** Hay 9 deciles que la dividen en 10 partes iguales: (primero al noveno decil).

**Percentiles:** Hay 99 percentiles que dividen a una serie en 100 partes iguales: (primero al noventa y nueve percentil).

```
> duration = faithful$eruptions      # the eruption durations
> quantile(duration, c(.32, .57, .98))
      32%      57%      98%
2.3952 4.1330 4.9330
```

Imagen ##. Ejemplo del cálculo de medidas de posición en R

<sup>29</sup> <http://www.monografias.com/trabajos14/medidasposicion/medidasposicion.shtml>  
<http://www.r-tutor.com/elementary-statistics/numerical-measures/percentile>

## Medidas de tendencia central<sup>30</sup>

Este tipo de medidas nos permiten identificar y ubicar el punto (valor) alrededor del cual se tienden a reunir los datos (“Punto central”). Estas medidas aplicadas a las características de las unidades de una muestra se les denomina estimadores o estadígrafos; mientras que aplicadas a poblaciones se les denomina parámetros o valores estadísticos de la población.

Al describir grupos de observaciones, con frecuencia es conveniente resumir la información con un solo número. Este número que, para tal fin, suele situarse hacia el centro de la distribución de datos se denomina medida o parámetro de tendencia central o de centralización.

El propósito de las medidas de tendencia central es:

- Mostrar en qué lugar se ubica la persona promedio o típica del grupo.
- Sirve como un método para comparar o interpretar cualquier puntaje en relación con el puntaje central o típico.
- Sirve como un método para comparar el puntaje obtenido por una misma persona en dos diferentes ocasiones.
- Sirve como un método para comparar los resultados medios obtenidos por dos o más grupos.

Las medidas de tendencia central más comunes son:

### Media

Es la medida de posición central más utilizada, la más conocida y la más sencilla de calcular, debido principalmente a que sus ecuaciones se prestan para el manejo algebraico, lo cual la hace de gran utilidad. Su principal desventaja radica en su sensibilidad al cambio de uno de sus valores o a los valores extremos demasiado grandes o pequeños. La media se define como la suma de todos los valores observados, dividido por el número total de observaciones.

### Mediana

Con esta medida podemos identificar el valor que se encuentra en el centro de los datos, es decir, nos permite conocer el valor que se encuentra exactamente en la mitad del conjunto de datos después que las observaciones se han ubicado en serie ordenada. Esta medida nos indica que la mitad de los datos se encuentran por debajo de este valor y la otra mitad por encima del mismo.

---

<sup>30</sup> <http://www.profesorenlinea.cl/matematica/EstadisticaMediaMedianaModa.htm>

<http://www.spssfree.com/curso-de-spss/analisis-descriptivo/media-mediana-moda-medidas-tendencia-central.html>

[http://es.wikipedia.org/wiki/Medidas\\_de\\_tendencia\\_central](http://es.wikipedia.org/wiki/Medidas_de_tendencia_central)

<http://www.cyclismo.org/tutorial/R/basicOps.html>

## Moda

La medida modal nos indica el valor que más veces se repite dentro de los datos; es decir, si tenemos la serie ordenada (2, 2, 5 y 7), el valor que más veces se repite es el número 2 quien sería la moda de los datos. Es posible que en algunas ocasiones se presente dos valores con la mayor frecuencia, lo cual se denomina Bimodal o en otros casos más de dos valores, lo que se conoce como multimodal.

En conclusión las Medidas de tendencia central, nos permiten identificar los valores más representativos de los datos, de acuerdo a la manera como se tienden a concentrar. La Media nos indica el promedio de los datos; es decir, nos informa el valor que obtendría cada uno de los individuos si se distribuyeran los valores en partes iguales. La Mediana por el contrario nos informa el valor que separa los datos en dos partes iguales, cada una de las cuales cuenta con el cincuenta por ciento de los datos. Por último la Moda nos indica el valor que más se repite dentro de los datos.

```
> tree$LFBM
 [1] 0.430 0.400 0.450 0.820 0.520 1.320 0.900 1.180 0.480 0.210 0.270 0.310
[13] 0.650 0.180 0.520 0.300 0.580 0.480 0.580 0.580 0.410 0.480 1.760 1.210
[25] 1.180 0.830 1.220 0.770 1.020 0.130 0.680 0.610 0.700 0.820 0.760 0.770
[37] 1.690 1.480 0.740 1.240 1.120 0.750 0.390 0.870 0.410 0.560 0.550 0.670
[49] 1.260 0.965 0.840 0.970 1.070 1.220
```

```
> mean(tree$LFBM)
 [1] 0.7649074
> median(tree$LFBM)
 [1] 0.72
> quantile(tree$LFBM)
   0%   25%   50%   75%  100%
0.1300 0.4800 0.7200 1.0075 1.7600
> min(tree$LFBM)
 [1] 0.13
> max(tree$LFBM)
 [1] 1.76
> var(tree$LFBM)
 [1] 0.1429382
> sd(tree$LFBM)
 [1] 0.3780717
```

Ejemplo del cálculo de medidas de tendencia central en R

## Medidas de dispersión<sup>31</sup>

Las Medidas de dispersión nos permiten reconocer que tanto se dispersan los datos alrededor del punto central; es decir, nos indican cuanto se desvían las observaciones alrededor de su promedio aritmético (Media). Este tipo de medidas son parámetros informativos que nos permiten conocer como los valores de los datos se reparten a través de eje X, mediante un valor numérico que representa el promedio de dispersión de los datos. Las medidas de dispersión más importantes y las más utilizadas son la Varianza y la Desviación estándar (o Típica).

Indican la mayor o menor concentración de los datos con respecto a las medidas de centralización. Los más usuales son: la desviación típica, el coeficiente de variación, el rango, la varianza...

### Varianza

Esta medida nos permite identificar la diferencia promedio que hay entre cada uno de los valores respecto a su punto central (Media). Este promedio es calculado, elevando cada una de las diferencias al cuadrado (Con el fin de eliminar los signos negativos), y calculando su promedio o media; es decir, sumado todos los cuadrados de las diferencias de cada valor respecto a la media y dividiendo este resultado por el número de observaciones que se tengan. Si la varianza es calculada a una población (Total de componentes de un conjunto).

### Desviación típica

Esta medida nos permite determinar el promedio aritmético de fluctuación de los datos respecto a su punto central o media. La desviación estándar nos da como resultado un valor numérico que representa el promedio de diferencia que hay entre los datos y la media.

### Coeficiente de variación

Cuando se desea hacer referencia a la relación entre el tamaño de la media y la variabilidad de la variable, se utiliza el coeficiente de variación.

Su fórmula expresa la desviación estándar como porcentaje de la media aritmética, mostrando una mejor interpretación porcentual del grado de variabilidad que la desviación típica o estándar. Por otro lado presenta problemas ya que a diferencia de la desviación típica este coeficiente es variable ante cambios de origen. Por ello es importante que todos los valores sean positivos y su media dé, por tanto, un valor positivo. A mayor valor del coeficiente de variación mayor heterogeneidad de los valores de la variable; y a menor C.V., mayor homogeneidad en los valores de la variable. Suele representarse por medio de las siglas C.V.

---

<sup>31</sup> <http://www.spssfree.com/curso-de-spss/analisis-descriptivo/varianza-desviacion-medidas-de-dispersion.html>  
[http://es.wikipedia.org/wiki/Coeficiente\\_de\\_variaci%C3%B3n](http://es.wikipedia.org/wiki/Coeficiente_de_variaci%C3%B3n)

```
# Dispersión
pesos <- rnorm( 1000, 3, 0.8 )
range( pesos )
```

```
## [1] 0.1402 5.9203
```

```
CV <- sd( pesos ) / mean( pesos )
CV
```

```
## [1] 0.2716
```

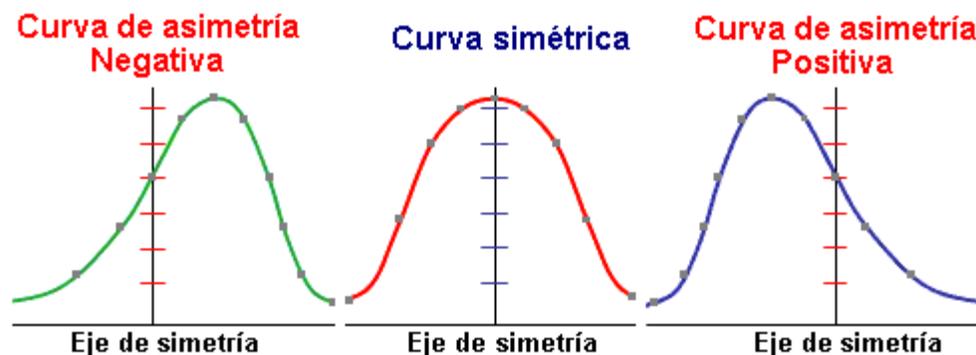
Ejemplo del cálculo de medidas de dispersión en R

### Medidas de distribución<sup>32</sup>

Las medidas de distribución nos permiten identificar la forma en que se separan o aglomeran los valores de acuerdo a su representación gráfica. Estas medidas describen la manera como los datos tienden a reunirse de acuerdo con la frecuencia con que se hallen dentro de la información. Su utilidad radica en la posibilidad de identificar las características de la distribución sin necesidad de generar el gráfico.

#### Asimetría

Esta medida nos permite identificar si los datos se distribuyen de forma uniforme alrededor del punto central (Media aritmética). La asimetría presenta tres estados diferentes, cada uno de los cuales define de forma concisa como están distribuidos los datos respecto al eje de asimetría. Se dice que la asimetría es positiva cuando la mayoría de los datos se encuentran por encima del valor de la media aritmética, la curva es Simétrica cuando se distribuyen aproximadamente la misma cantidad de valores en ambos lados de la media y se conoce como asimetría negativa cuando la mayor cantidad de datos se aglomeran en los valores menores que la media.



Tres estados de la Asimetría

#### Curtosis o apuntamiento

<sup>32</sup> <http://www.spssfree.com/curso-de-spss/analisis-descriptivo/medidas-de-distribucion-curtosis-asimetria.html>  
[http://cran.r-project.org/doc/contrib/Chicana-Introduccion\\_al\\_uso\\_de\\_R.pdf](http://cran.r-project.org/doc/contrib/Chicana-Introduccion_al_uso_de_R.pdf)

Esta medida determina el grado de concentración que presentan los valores en la región central de la distribución. Por medio del Coeficiente de Curtosis, podemos identificar si existe una gran concentración de valores (Leptocúrtica), una concentración normal (Mesocúrtica) ó una baja concentración (Platicúrtica).

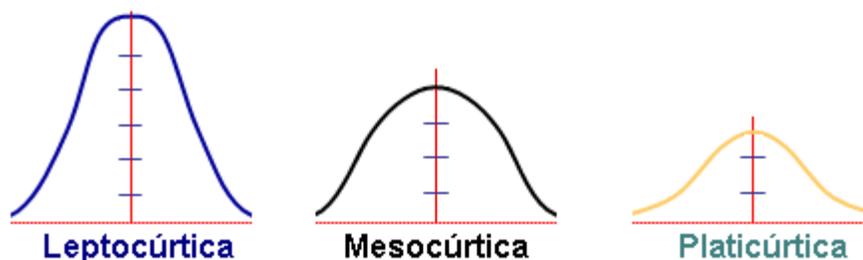


Imagen ##. Tres tipos de concentraciones de la Curtosis

```
- Sesgo  
> skewness(MMC1)  
[1] 0.0005004876  
  
- Curtosis  
> kurtosis(MMC1)  
[1] -0.4139902
```

Ejemplo del cálculo de medidas de distribución en R

### Curva Normal<sup>33</sup>

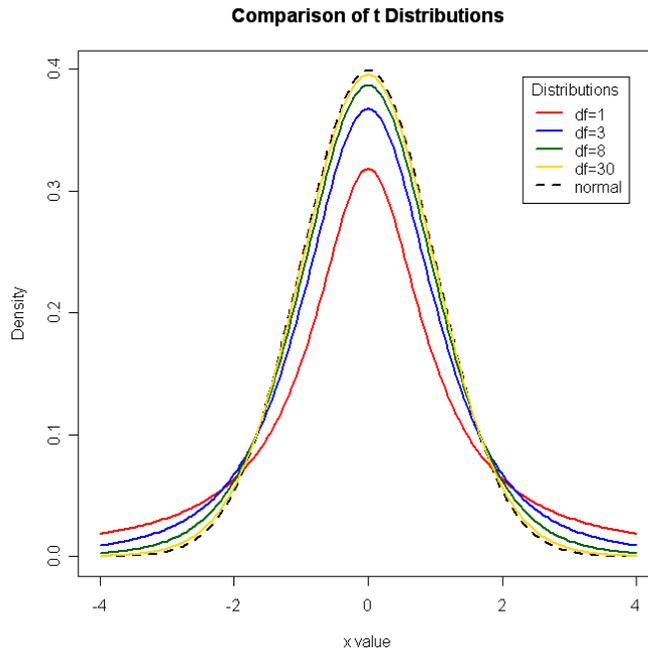
La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos. Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.

La distribución normal también es importante por su relación con la estimación por mínimos cuadrados, uno de los métodos de estimación más simples y antiguos.

Algunos ejemplos de variables asociadas a fenómenos naturales que siguen el modelo de la normal son:

- caracteres morfológicos de individuos como la estatura;
- caracteres fisiológicos como el efecto de un fármaco;
- caracteres sociológicos como el consumo de cierto producto por un mismo grupo de individuos;
- caracteres psicológicos como el cociente intelectual;
- nivel de ruido en telecomunicaciones;
- errores cometidos al medir ciertas magnitudes; etc.

<sup>33</sup> [http://es.wikipedia.org/wiki/Distribuci%C3%B3n\\_normal](http://es.wikipedia.org/wiki/Distribuci%C3%B3n_normal)



Ejemplo del cálculo la curva normal en R

### Diagramas de cajas<sup>34</sup>

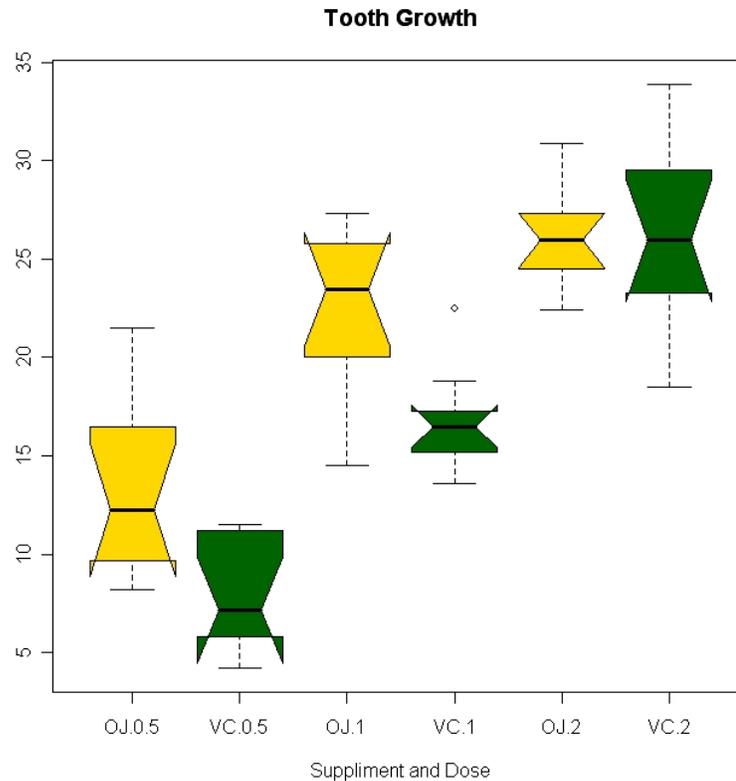
Los diagramas de Caja-Bigotes (boxplots o box and whiskers) son una presentación visual que describe varias características importantes, al mismo tiempo, tales como la dispersión y simetría.

Para su realización se representan los tres cuartiles y los valores mínimo y máximo de los datos, sobre un rectángulo, alineado horizontal o verticalmente.

Una gráfica de este tipo consiste en una caja rectangular, donde los lados más largos muestran el recorrido intercuartílico. Este rectángulo está dividido por un segmento vertical que indica donde se posiciona la mediana y por lo tanto su relación con los cuartiles primero y tercero (recordemos que el segundo cuartil coincide con la mediana).

Esta caja se ubica a escala sobre un segmento que tiene como extremos los valores mínimo y máximo de la variable. Las líneas que sobresalen de la caja se llaman bigotes. Estos bigotes tienen un límite de prolongación, de modo que cualquier dato o caso que no se encuentre dentro de este rango es marcado e identificado individualmente.

<sup>34</sup> <http://www.estadisticaparatodos.es/taller/graficas/cajas.html>

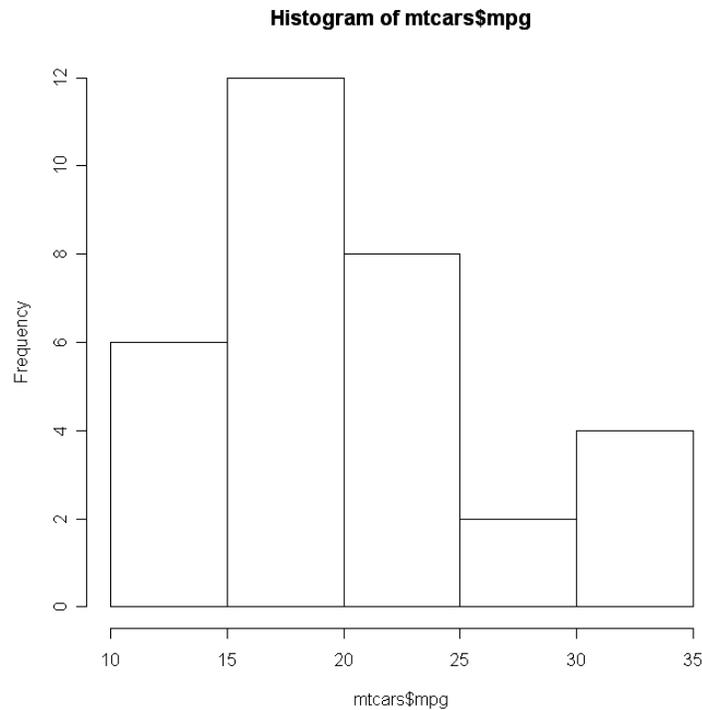


Ejemplo del diagrama de caja y bigotes en R

### Histograma<sup>35</sup>

Un Histograma es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados, ya sea en forma diferencial o acumulada. Sirven para obtener una "primera vista" general, o panorama, de la distribución de la población, o la muestra, respecto a una característica, cuantitativa y continua, de la misma y que es de interés para el observador (como la longitud o la masa). De esta manera ofrece una visión en grupo permitiendo observar una preferencia, o tendencia, por parte de la muestra o población por ubicarse hacia una determinada región de valores dentro del espectro de valores posibles (sean infinitos o no) que pueda adquirir la característica. Así pues, podemos evidenciar comportamientos, observar el grado de homogeneidad, acuerdo o concisión entre los valores de todas las partes que componen la población o la muestra, o, en contraposición, poder observar el grado de variabilidad, y por ende, la dispersión de todos los valores que toman las partes, también es posible no evidenciar ninguna tendencia y obtener que cada miembro de la población toma por su lado y adquiere un valor de la característica aleatoriamente sin mostrar ninguna preferencia o tendencia, entre otras cosas.

<sup>35</sup> <http://es.wikipedia.org/wiki/Histograma>



Ejemplo del histograma en R

### Correlación<sup>36</sup>

La correlación indica la fuerza y el sentido de una relación lineal y proporcionalidad entre dos variables estadísticas. Se considera que dos variables cuantitativas están correlacionadas cuando los valores de una de ellas varían sistemáticamente con respecto a los valores homónimos de la otra: si tenemos dos variables (A y B) existe correlación si al aumentar los valores de A lo hacen también los de B y viceversa. La correlación entre dos variables no implica, por sí misma, ninguna relación de causalidad.

La correlación estadística constituye una técnica estadística que nos indica si dos variables están relacionadas o no. La correlación es una herramienta poderosa que brinda piezas vitales de información.

Los principales componentes elementales de una correlación, son la fuerza y el sentido:

- La fuerza extrema según el caso, indica que la relación es fuerte o es débil.
- El sentido mide la variación de los valores de B con respecto a A: si al crecer los valores de A lo hacen los de B, la relación es directa; si al crecer los valores de A disminuyen los de B, la relación es inversa.

Entre las formas de representar la correlación están:

---

<sup>36</sup> <http://es.wikipedia.org/wiki/Correlaci%C3%B3n>

## Matriz de Correlaciones

Presenta numéricamente la correlación existente entre las diferentes variables expresada con valores entre -1 y 1.

```
> showData(Mexico, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80, maxheight=30)

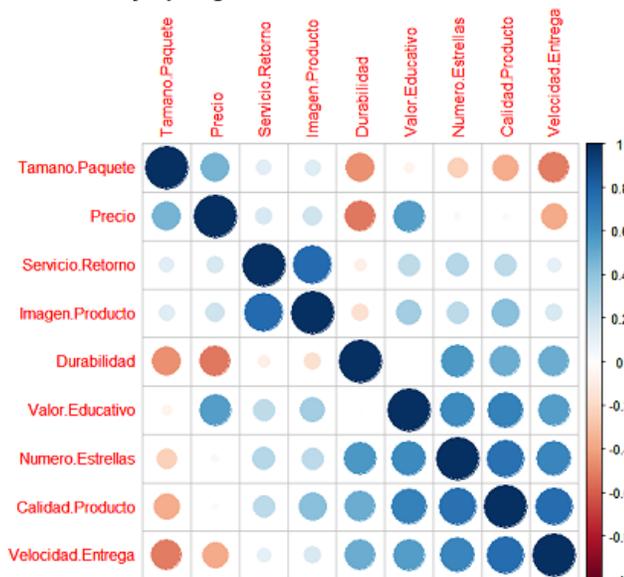
> cor(Mexico[,c("Costa.Rica", "El.Salvador", "Guatemala", "Honduras", "Nicaragua", "Panama")],
+ use="complete")
      Costa.Rica El.Salvador Guatemala Honduras Nicaragua Panama
Costa.Rica  1.00000000 -0.02576306 -0.07793314 -0.096323215  0.519987812  0.4050191
El.Salvador -0.02576306  1.00000000  0.66988792  0.644137475  0.311254912  0.8415630
Guatemala  -0.07793314  0.66988792  1.00000000  0.538797492  0.265223356  0.5852378
Honduras    -0.09632321  0.64413747  0.53879749  1.000000000  -0.006031797  0.3079463
Nicaragua   0.51998781  0.31125491  0.26522336  -0.006031797  1.000000000  0.6445250
Panama      0.40501910  0.84156298  0.58523779  0.307946346  0.644525026  1.0000000
```

Ejemplo de la matriz de correlación en R

Mapa de correlaciones

Presenta gráficamente mediante círculos y colores la correlación que existe entre las diferentes variables. El tamaño del círculo representa el sentido, mientras más grande es el círculo mayor es la correlación tanto en sentido positivo, como negativo. El color representa la fuerza, mientras más oscuro, ya sea azul o rojo, más fuerza de correlación hay entre las variables tanto en sentido positivo como negativo.

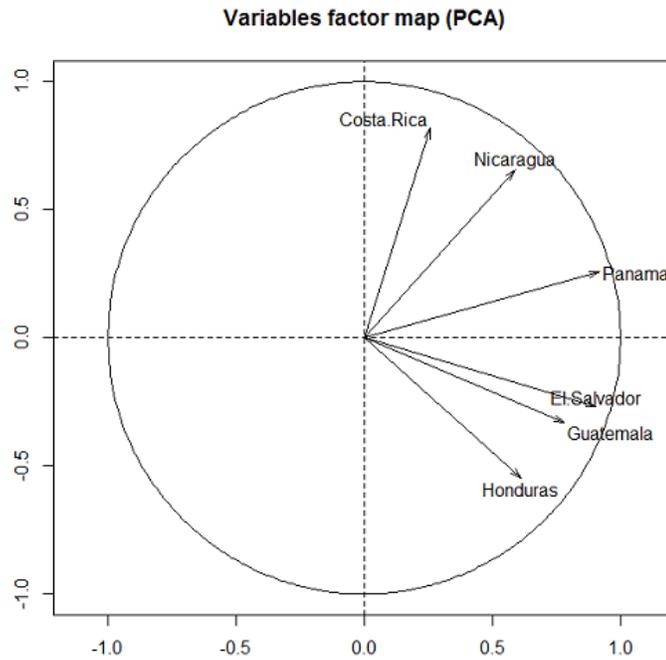
Correlación EjemploAlgoritmosRecomendacion.csv usando Pearson



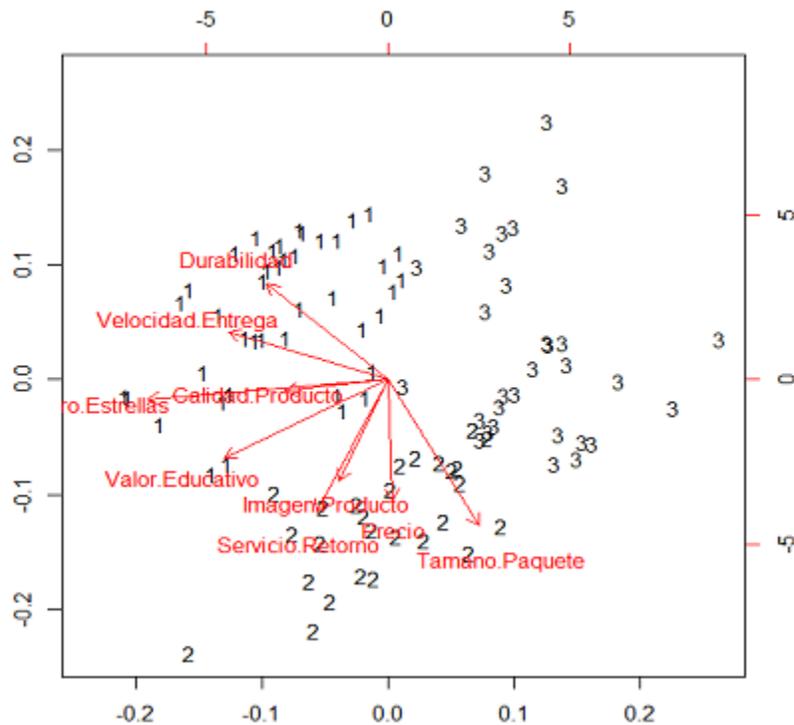
Ejemplo del mapa de correlaciones en R

## Círculo de correlaciones

Presenta mediante una serie de líneas rectas la correlación que existe entre las diferentes variables. La longitud de la línea representa la fuerza de la correlación y la dirección de la línea el sentido. Puede darse el caso en el que las líneas se dirijan en sentidos opuestos, lo indica una mayor correlación negativa. También es posible representar dentro de la serie de líneas los conjuntos formados por los valores y poder reconocer grupos que presentan similitudes en sus variables.



Ejemplo del círculo de correlaciones en R



Ejemplo del círculo de correlaciones con el conjunto agrupado de datos en R

### Análisis de Conglomerados<sup>37</sup>

El Análisis clúster (AC), también llamado en ocasiones Análisis de conglomerados, se dice habitualmente que es una técnica estadística clasificadora, pero, en realidad, es una técnica que, como el Análisis de componentes principales (ACP) o como el Análisis factorial (AF), pretende representar una realidad que no conseguimos visualizar, una realidad cuya representación original es multidimensional y es imposible que la podamos ver en su estado puro.

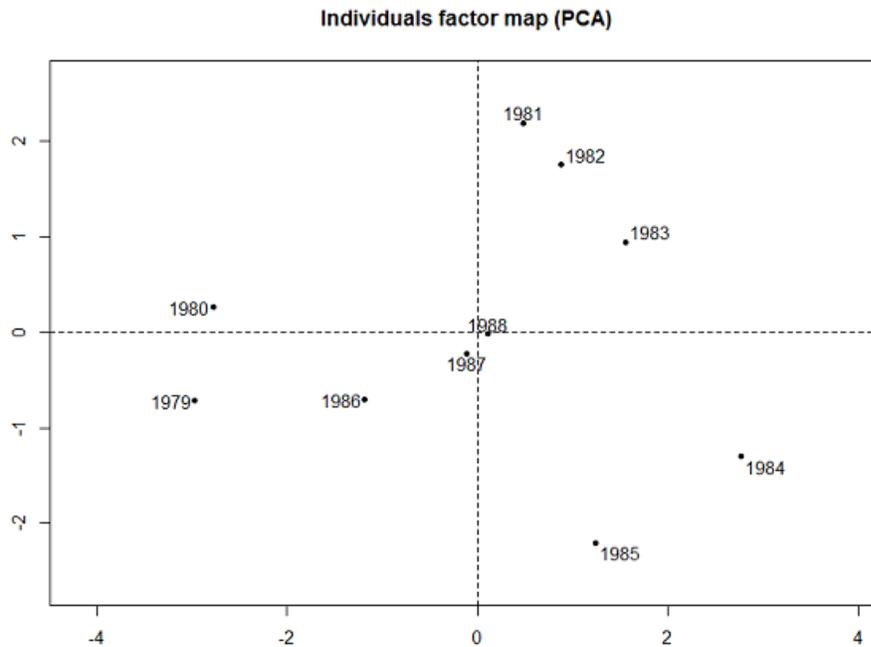
En el fondo tanto ACP, como AF, como AC son técnicas que tratan de representar una nube de puntos original situada en un espacio de tantas dimensiones que es imposible visualizar. Y cada una de ellas, también, en el fondo, puede ser usada como método clasificatorio, como método para crear subpoblaciones, subgrupos.

La diferencia fundamental entre ellas es la forma de presentación que utilizan, la forma de resolver el problema de no visualización de la nube de puntos originales. El ACP y el AF lo hacen construyendo una nube de puntos de la misma naturaleza pero de menor número de dimensiones perdiendo una parte de la información original. Sin embargo, el AC lo que hace es crear una representación distinta a la de la nube de puntos. Crea otro tipo de representación. Cambia la forma: no lo hace mediante una nube de puntos, lo hace mediante un dendograma.

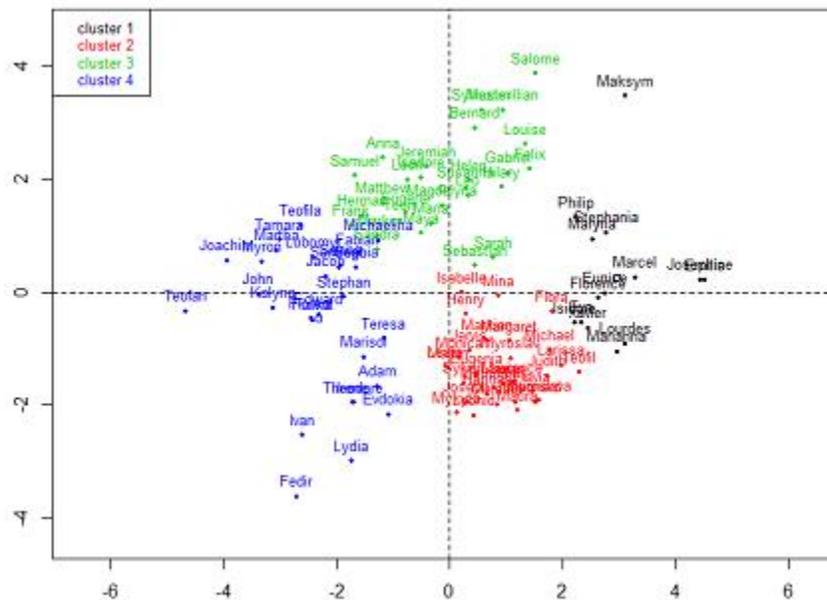
Cada una de las opciones tiene sus ventajas y sus desventajas, como iremos viendo a continuación.

<sup>37</sup> <https://estadisticaorquestainstrumento.wordpress.com/2013/01/02/tema-19-analisis-cluster/>

El ACP y el AF respetan el tipo de representación: una nube de puntos, pero al reducir dimensiones se pierde información y esto es un problema, especialmente si la pérdida es importante. El AC respeta la nube de puntos original, no reduce dimensiones y, por lo tanto, no se pierde información, pero, eso sí, se cambia el mecanismo de representación, se cambia el estilo de representación. Representamos la nube de puntos mediante un dendrograma. Digamos que en el ACP y el AF se hace una representación figurativa y en el AC se hace una representación abstracta.

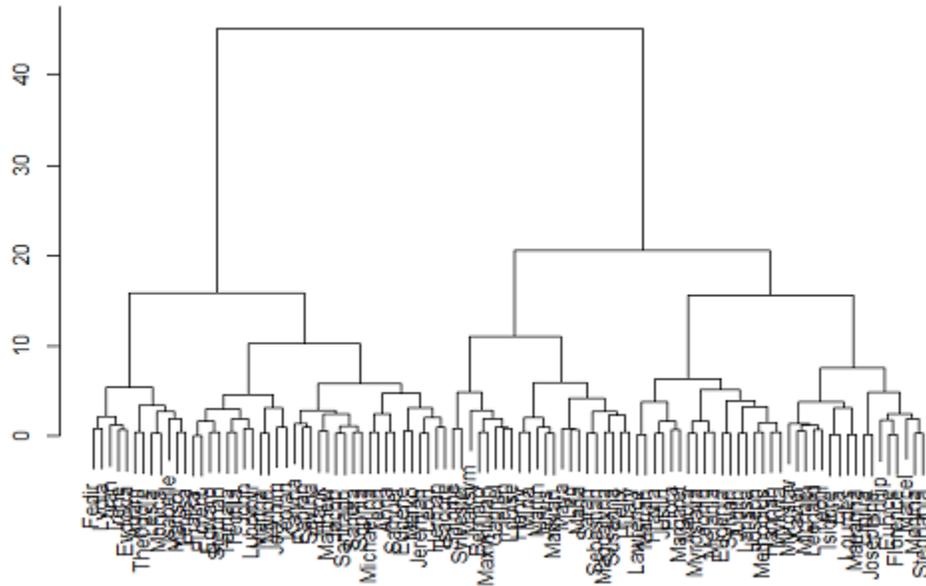


Ejemplo del grafico de Análisis ACP



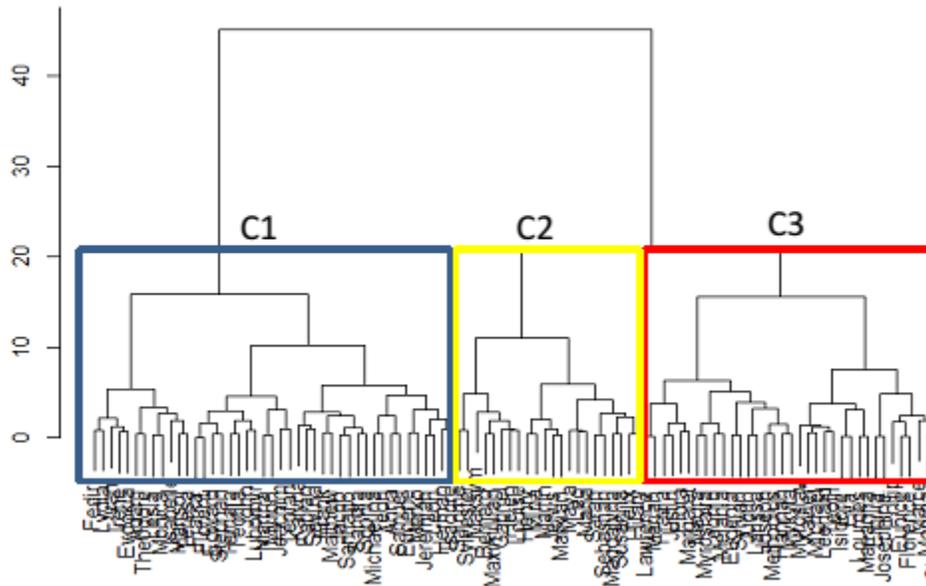
Ejemplo del grafico de conglomerados en R

**Cluster Dendrogram for Solution HClust.1**



Ejemplo del grafico del dendograma en R

**Cluster Dendrogram for Solution HClust.1**



Ejemplo del grafico del dendograma con aglomerados marcados en R

**Centros de Gravedad**

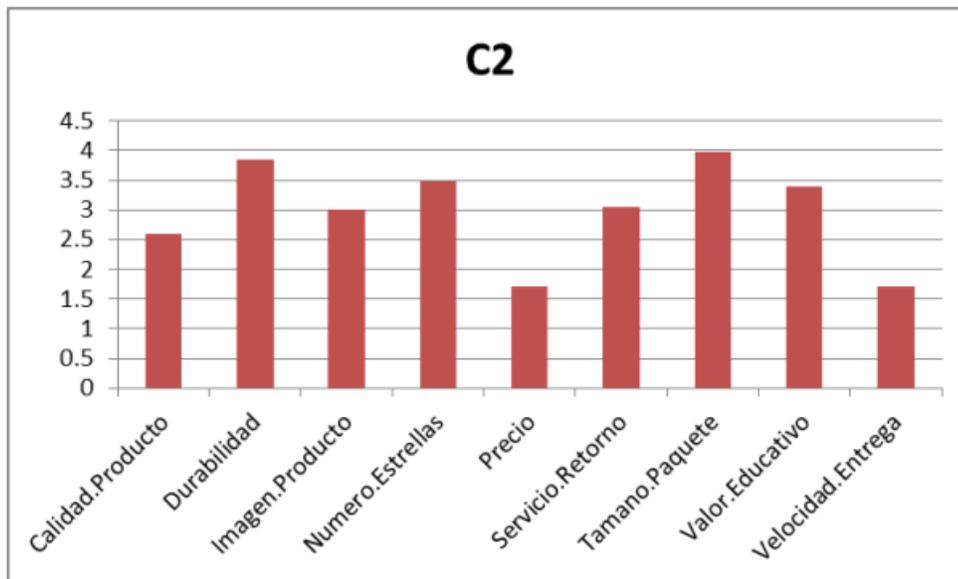
Representa el centro de la nube de puntos que representan un conglomerado de valores.

Hay dos tipos:

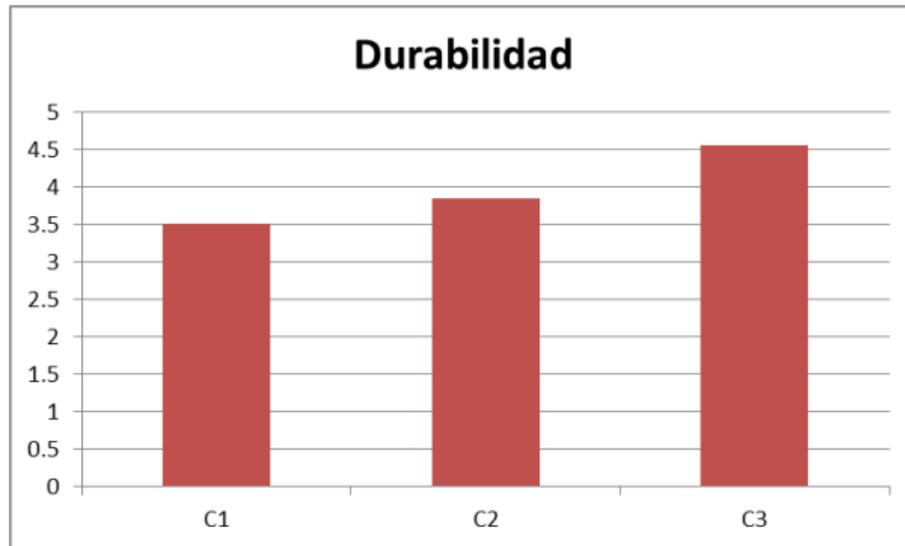
- Horizontal: Compara los valores obtenidos por cada conglomerado para cada una de las variables en estudio.
- Vertical: Compran todos los valores obtenidos de todas las variables para un conglomerado en particular.

	Calidad.Producto	Durabilidad	Imagen.Producto	Numero.Estrellas	Precio	Servicio.Retorno	Tamano.Paquete	Valor.Educativo	Velocidad.Entrega
C1	2.019767	3.503488	2.496512	2.288372	1.151163	2.516279	3.748837	2.35814	1.262791
C2	2.590909	3.843182	3.006818	3.472727	1.709091	3.040909	3.979545	3.390909	1.7
C3	2.7057143	4.5571429	2.54	3.8914286	0.8885714	2.6114286	2.8514286	3.3028571	2.4014286

Ejemplo de la tabla de centros de gravedad



Ejemplo del grafico de centros de gravedad vertical



Ejemplo del grafico de centros de gravedad horizontal

### Estadística Inferencial

La Estadística Inferencial es una amplia rama de la estadística que comprende la deducción y el análisis de información que, a partir de un conjunto de datos llamado muestra, pretende, dado el conocimiento de la distribución de los datos (particularmente la Distribución Normal) y de un conjunto de estimadores y estadísticos, obtener resultados que se acerquen a lo que se puede apreciar en el universo total de dicha variable, mejor conocida como población.

La Estadística Inferencial se puede clasificar en dos corrientes teóricas: La Estadística Clásica y la Estadística Bayesiana.

La Estimación Estadística Paramétrica Clásica plantea tres tipos de problemas:

**Estimación puntual:** Con la cual se busca dar un valor a cierto parámetro hacia el cual se desea estimar o inferir.

**Estimación por intervalos:** En la cual se construyen Intervalos de Confianza dentro de los cuales puede darse el valor del parámetro poblacional.

**Pruebas de hipótesis:** Con las cuales se busca contrastar información acerca del parámetro dados unos valores hipotéticos.

En resumen, la forma en que trabaja la Estadística Inferencial se da, habitualmente, de la siguiente manera: Se tiene un experimento  $X$ , el cual es repetido varias veces y mediante el cual se obtiene una muestra compuesta por Variables Aleatorias Independientes idénticamente Distribuidas, cuya función de Distribución también es conocida. Cualquier función de la muestra que no dependa del parámetro a estimar es un Estadístico y aquel estadístico que se utiliza para inferir sobre el parámetro desconocido es un Estimador. Los ejemplos más comunes de estadísticos son el Total Muestral, la Media Muestral, la Varianza Muestral, los estadísticos de orden, entre otros.

Dados estos conceptos, se asume en principio que los datos relacionados con el experimento X siguen una Distribución Normal, lo cual es un supuesto necesario y suficiente para comenzar a realizar inferencia Paramétrica. Para que esto sea posible, es necesario entonces definir unas distribuciones de muestreo que se vinculen a la Distribución Normal y que por tanto, construyan estimaciones confiables.

Función en R para estimaciones puntuales, por intervalos y prueba de hipótesis.

```
>t.test (datos_x, datos_y =NULL, alternative = "two.sided", mu = 0, paired = FALSE,  
var.equal = FALSE, conf.level= 0.95)
```

### **Modelos de Regresión Lineal en R**

En general, los modelos de regresión lineal son modelos probabilísticos basados en una función lineal, mediante la cual se expresa una variable (conocida como variable endógena o explicada), en función de otro conjunto de variables (llamadas exógenas o explicativas).

En resumen, se podrían considerar los siguientes pasos como los básicos a seguir en la construcción de un modelo de regresión lineal

Definir el modelo matemático a estudiar con sus respectivas hipótesis.

Estimación de los parámetros del modelo.

Inferencias sobre los parámetros del modelo.

Diagnóstico y conclusiones sobre el modelo.

La función más importante para la elaboración del análisis de Regresión Lineal en R es la función `lm()`, la cual abrevia la frase "lineal model" (Modelo Lineal). Esta función tiene unas características muy importantes, como por ejemplo que no ofrece ningún resultado por pantalla sino que simplemente crea un objeto que precisamente será un modelo de regresión lineal.

Este objeto puede ser referenciado por cualquier otra función para realizar algunos otros procedimientos como por ejemplo la construcción de un modelo autoregresivo o de un análisis de varianza. En general, la función `lm()` tiene la siguiente estructura:

```
>lm(formula, data, subset, weights, method = "qr", model = TRUE, x = FALSE, y = FALSE,  
qr = TRUE, contrasts = NULL, ...)
```

## ANEXO 2

# SISTEMAS DE DESARROLLO Y PRODUCCIÓN DEL MINISTERIO DE SALUD

No.	Sistema	Lenguaje de programación	Base de datos	Modulos o subsistemas	Objetivo	Areas o dependencias que lo utilizan
1	De Monitoreo, Evaluación y Vigilancia Epidemiológica del VIH/SIDA (SUMEVE)	PHP	MySQL	No	Control y seguimiento de la evolución de la epidemia de SIDA en el Salvador	Nivel Central, Regional, Local y otras instituciones
2	De Morbi-Mortalidad	PHP	MySQL	No	Registrar egresos hospitalarios, las consultas curativas, atenciones preventivas para el monitoreo y seguimiento de las patologías, defunciones de registro de alcaldía.	Nivel Central, Regional y Local
3	De Lesiones de Causa Externa (SILEX)	PHP	MySQL	No	Registrar lesiones de causa externa para el monitoreo y seguimiento de las patologías.	Nivel Central, Regional y Local
4	De Vigilancia Epidemiológica (VIGEPES)	PHP	MySQL	No	Vigilancia Epidemiológica semanal, eventos de notificación inmediata y resultados de laboratorio (diagnósticos vigilancia epidemiológica)	Nivel Central, Regional, Local y otras instituciones

No.	Sistema	Lenguaje de programación	Base de datos	Modulos o subsistemas	Objetivo	Areas o dependencias que lo utilizan
5	Sistema de Fichas Familiares	PHP	POSTGRE SQL	No	Proporcionar una herramienta de captura y administración de la Ficha Familiar que permita obtener información oportuna referente al diagnóstico de salud de las familias y comunidades, de las áreas intervenidas en el nuevo modelo de salud familiar por los Equipos Comunitarios de Salud Familiar.	Nivel Central del MSPAS, Regiones de Salud y Unidad de Salud
6	Sistema de Encuestas del Ministerio de Salud.	PHP	MySQL	No	Sistema para la captura procesamiento y análisis estadístico básico de diferentes iniciativas de las direcciones, unidades y programas del Ministerio de Salud.	Nivel Central del MSPAS, Regiones de Salud y Hospitales
7	Sistema de Información Gerencial	PHP	POSTGRE SQL	No	Crear Indicadores para la toma de decisiones	Jefaturas y Publico en General
8	Sistema de Fondos Externos	PHP	POSTGRE SQL	No	Crear una herramienta que sirva de apoyo en la gestión de proyectos de donación realizados en la Unidad de Fondos Externos del Ministerio de Salud Pública de El Salvador.	Fondos Externos y dependencias del nivel central.

No.	Sistema	Lenguaje de programación	Base de datos	Modulos o subsistemas	Objetivo	Areas o dependencias que lo utilizan
9	Sistema Integral de Atención al Paciente (SIAP)	PHP	MySQL	Modulo de Identificación del Paciente	Manejo de información de los expedientes clínicos de los pacientes, en las áreas de emergencias, consulta externa , farmacia, servicios de apoyo y citas médicas	Hospital Nacional Rosales, Cojutepeque, San Vicente, Usulután, Santa Ana, Metapan, San Rafael, Nva.Concep, Chalatenango, Sta. R. lima, La Unión, Nva.Guadalup, Zacatecoluca, U.de S. Díaz del Pinal, U d S Ciudad Mujer
				Modulo de Citas médicas	permite generar, controlar y compartir toda la información relacionada al otorgamiento de citas. Para así poder agilizar el funcionamiento del área de consulta externa y las entidades involucradas.	Hospital Nacional Rosales, San Rafael y Santa Ana
			MySQL	Modulo de Seguimiento clínico	Este modulo permitirá registrar y consultar información del seguimiento del expediente clínico de un paciente ingresado al sistema.	Hospital Nacional Rosales
				Modulo de Farmacia	Este modulo tiene la finalidad de facilitar el proceso de despacho de medicamentos, recepción y procesamiento de solicitudes de medicamentos de la Consulta Externa, Control de despacho de recetas diarias y repetitivas, Control de existencia de medicamentos.	Hospital Nacional Rosales

No.	Sistema	Lenguaje de programación	Base de datos	Modulos o subsistemas	Objetivo	Areas o dependencias que lo utilizan
10	Sistema Integral de Atención al Paciente (SIAP)	PHP	MySQL	Modulo de Laboratorio clínico	Este modulo tiene la finalidad de facilitar el ingreso y consulta de información de forma oportuna, permitiendo llevar el control de citas a pacientes que solicitan el servicio de laboratorio clínico, interfaces de captura de los resultados de exámenes por paciente para que posteriormente puedan ser consultados por el médico en el sistema en línea.	Hospital Nacional Rosales
11	Estadístico de Producción de Servicios SEPS	PHP	Mysql	No	Medir la producción de los servicios de salud que brinda el MSPAS	Unidad de Información
12	Sistema de Correspondencia SCC	PHP	Mysql	No	Sistema para el control y seguimiento de la correspondencia.	Nivel Central del MSPAS
13	Sistema de Inventario de Equipo OCS / GLPI	PHP	Mysql	No	Sistema para el control y administración del inventario técnico del equipo de cómputo.	Nivel Central del MSPAS, Regiones de Salud y Unidad de Salud
14	Sistema Helpdesk GLPI	PHP	Mysql	No	Llevar el control de incidencias reportadas en el área de Soporte Técnico	-

No.	Sistema	Lenguaje de programación	Base de datos	Modulos o subsistemas	Objetivo	Areas o dependencias que lo utilizan
15	Centro Virtual de Documentación Regulatoria	ASP	Mysql	No	Brindar acceso al público y establecimientos del sector salud de las diferentes leyes, normas, y demás documentos regulatorios emitidos por el MINSAL	Nivel Central del MSPAS, Regiones de Salud y Unidad de Salud, acceso público
16	De Activo Fijo (SIAF)	.NET	SQL Server	1. Administración 2. Técnicos	Administración de los activos fijos institucionales a nivel nacional	Unidad de Patrimonio de todos los establecimientos del MINSAL a nivel nacional
17	De Control de Servicio Social (SCSS)	.NET	SQL Server	1. Administración 2. Técnicos	Controlar el servicio social brindado por los estudiantes egresados de las instituciones formadoras de profesionales en salud (Medicina, Odontología, Enfermería y Técnicos en Salud) que optan por plazas de servicio social en los distintos establecimientos.	Recursos Humanos, Nivel Central como administrador del sistema y las cinco regiones a nivel nacional
18	Nacional de Abastecimiento (SINAB)	.NET	SQL Server	1. Administración	1. Administrar los siguientes catálogos: a. Listado de Productos b. Listado de Dependencias c. Listado de Establecimientos d. Listado de Proveedores 2. Creación de Usuarios y Roles	Unidad de Abastecimiento

## ANEXO 3

# RESUMEN DE INDICADORES PRIORIZADOS Y ESTANDARIZADOS EN SALUD

Luego de la propuesta del Consejo Directivo de la Organización Panamericana de la Salud en el año 2000, se creó en nuestro país una Guía de Indicadores Priorizados y Estandarizados en Salud.

Los indicadores de salud se dividen en dos amplias categorías: Indicadores epidemiológicos e indicadores operacionales.

Los indicadores epidemiológicos se utilizan para estimar la magnitud, trascendencia y vulnerabilidad de los diferentes grupos de población ante una situación de salud determinada, en un espacio y tiempo definido. Así mismo, los indicadores epidemiológicos permiten conocer la incidencia y prevalencia de enfermedades sujetas a vigilancia, así como la mortalidad de la población.

Los indicadores operacionales miden el trabajo realizado, ya sea en función de la cantidad o de la calidad de ellos, además, miden la cantidad de actividades y procedimientos realizados.

Los indicadores operacionales pueden ser de tres tipos: de productividad, costo promedio y calidad técnica.

Las medidas de frecuencias utilizadas en cualquier tipo de indicador son las siguientes:

**Cifras absolutas:** Es la expresión simple de la cantidad, da una idea de la magnitud o volumen real de un suceso. Tiene utilidad para la asignación de recursos de camas, personal y recursos físicos necesarios para satisfacer la demanda. El uso de cifras absolutas tiene limitaciones, puesto que no aluden a la población de la cual se obtienen.

**Razón:** Es el cociente de dos variables, los valores del numerador y del denominador son independientes, ninguno está contenido en el otro.

**Proporción:** Es el cociente de dos variables, el numerador está contenido en el denominador.

**Porcentaje:** La gran mayoría de los indicadores son proporciones para facilitar su interpretación, se ha preferido expresarlos como porcentajes, ejemplo: “cobertura” se refiere al porcentaje de una población que recibe atención en salud. Expresa una cantidad como un número de partes por cien unidades.

**Taza:** Es la magnitud del cambio de la variable que mide un evento por unidad de cambio de otra (el tiempo), en relación al tamaño de la población que se encuentra en riesgo de presentar el evento.

Los indicadores priorizado los dividiremos en una categorización más específica a continuación:

- Indicadores demográficos

Ejemplos.

<p><b>Porcentaje de población Masculina</b></p>	$\frac{\text{Total de población masculina}}{\text{Total de población ambos sexos}} \times 100$	<p>Se expresa como porcentaje correspondiente al sexo masculino en relación al cien por ciento de la población total, en un punto de tiempo y área geográfica específica. <b>Periodicidad:</b> anual</p>
<p><b>Tasa de abortos</b></p>	$\frac{\text{Numero de abortos}}{\text{Numero de mujeres de 10 a 49 años}} \times 100$	<p>La tasa de abortos es el número estimado de abortos por 1,000 mujeres de 10 a 49 años de edad en un determinado año. Esta tasa no debe confundirse con la razón de abortos.</p>
<p><b>Tasa de mortalidad general</b></p>	$\frac{\text{Numero de defunciones todas causas}}{\text{Población total}} \times 100$	<p>La tasa de mortalidad (llamada también tasa bruta de mortalidad) es el número de muertes que ocurren en un período de tiempo, por todas las causas, dividido por la población expuesta al riesgo de muerte durante el mismo</p>

período en estudio.  
Identifica la proporción de personas que fallecen por cualquier causa y a cualquier edad y sexo en una población.

- Indicadores de Infraestructura
  - Hospitales de tercer nivel
  - Hospitales de segundo nivel
  - Establecimientos de Primer nivel:
    - Unidades de Salud
    - Casas de Salud
    - Centros rural de nutrición
    - Clínica para empleados del MSPAS

- Indicadores Sociales.  
Ejemplos:

**Tasa de alfabetización**

Proporción de la población adulta de 15 y más años de edad que es letrada, expresada como porcentaje de la correspondiente población, para un determinado sexo, en un determinado país, territorio o área geográfica, en un punto de tiempo específico, usualmente a mitad de año.

**Proporción de población bajo la línea nacional de pobreza**

Porcentaje de la población viviendo bajo la línea nacional de pobreza, para un determinado país, territorio o área geográfica, en un periodo de tiempo específico, usualmente un año.

- Indicadores Económicos.  
Ejemplos:

**Gastos per cápita en salud**

$$\frac{\text{Gasto Nacional en Salud en el año}}{\text{Población del año}} \times 10,000$$

Gastos por persona al año en concepto de salud.

Otra categoría de indicadores. A ejemplo se mostraron los indicadores anteriores. Si quiere una explicación más detallada por cada indicador debe consultarse la “Guía de Indicadores Priorizados del Ministerio de Salud”.

- Indicadores Financieros
- Indicadores Recursos Humanos
- Indicadores de Rendimiento del Recurso Humano
- Indicadores Epidemiológicos