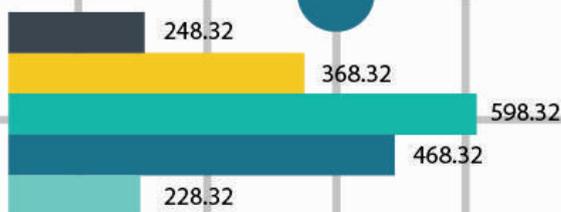




Metodología de investigación con contenidos estadísticos para el programa de Ingeniería de Sistemas de la Universidad Simón Bolívar



José Gerardo **Chacón Rangel**
Nahid Antuan **Bautista Vega**

Metodología de investigación con contenidos estadísticos para el programa de ingeniería de sistemas de la Universidad Simón Bolívar

**Metodología de investigación
con contenidos estadísticos
para el programa de ingeniería de sistemas
de la Universidad Simón Bolívar**



José Gerardo Chacón Rangel,
Nahid Antuan Bautista Vega

Metodología de investigación con contenidos estadísticos para el programa de ingeniería de sistemas de la Universidad Simón Bolívar

Grupos de investigación: Altos Estudios de Frontera (ALEF). Universidad Simón Bolívar, Cúcuta, Colombia.

Libro producto del proyecto de investigación titulado: Sistema informático fundamentado en técnicas de inteligencia artificial realizado sobre un microcomputador para el diagnóstico biomédico de pacientes hepáticos.

Fecha de inicio: marzo de 2015. *Fecha de finalización:* marzo de 2016.

ISBN 978-980-402-232-6
Depósito legal ZU2017000110

©José Gerardo Chacón Rangel

©Nahid Antuan Bautista Vega

Marzo de 2017 ~ Maracaibo, República Bolivariana de Venezuela

Todos los derechos reservados. Ninguna parte de esta publicación puede ser reproducida, almacenada en sistema recuperable o transmitida en medios electrónicos, mecánicos, fotocopias, grabaciones u otros, sin la previa autorización por escrito de la Universidad del Zulia. Los conceptos expresados en este documento son responsabilidad exclusiva del/os autor/es. Esta obra cumple con el requisito de evaluación por dos pares de expertos y depósito legal.

Diseño e ilustración de portada y contraportada: Eddy Enrique Barrios Boscán.

Diagramación e impresión:

Ediciones Astro Data, S.A. (edicionesastrodata@gmail.com)

Maracaibo, República Bolivariana de Venezuela

Metodología de investigación con contenidos estadísticos para el Programa de Ingeniería de Sistemas de la Universidad Simón Bolívar. / Chacón Rangel, J. G.; Bautista Vega, N. A. Universidad del Zulia. Maracaibo, Estado Zulia, República Bolivariana de Venezuela .2017. 205 p. 15,5 x 22,5 cm.

ISBN 978-980-402-232-6 / Depósito legal ZU2017000110

Palabras clave: Sistemas viables, cibernética, sistema informático biomédico.

Clasificación DEWEY: 600. 2017. 1a ed.

Universidad del Zulia - Sistema de Bibliotecas.

Publicaciones Científicas Universidad del Zulia:

www.luz.edu.ve ~ www.fc.es.luz.edu.ve / mfucchi@dri.luz.edu.ve

<http://www.fc.es.luz.edu.ve/cee/index.php/principios-del-cee/1-estructura-organizativa>

Yorberth Yannedy Montes de Oca Rojas (montesdeoca_yorberth@fc.es.luz.edu) +582614127639

Centro de Estudios de la Empresa, Edif. de Investigaciones Económicas "Gastón Parra Luzardo".

Facultad de Ciencias Económicas y Sociales. Núcleo Humanístico. Ciudad Universitaria. Av. 16 (Guajira).

Apartado de Correo 1540. Maracaibo. Estado Zulia. República Bolivariana de Venezuela.



Universidad del Zulia

Rector

PhD. Jorge Palencia Piña

Vicerrectora Académica

PhD. Judith Aular de Durán

Vicerrector Administrativo

PhD. Jesús Salón Crespo

Secretaria

PhD. Marlene Primera Galué

Coordinador-Secretario

*del Consejo de Desarrollo Científico,
Humanístico y Tecnológico (CONDES)*

PhD. Gilberto Vizcaíno

Comité Académico y Científico

PhD. Andrés Valdez Zepeda
(Universidad de Nuevo México, USA)

PhD. María Luisa Montánchez Torres
(Universidad Nacional de Educación UNAE, Ecuador)

PhD. Cecilia Socorro
(Universidad del Zulia, Venezuela)

PhD. Johann Enrique Pirela Morillo
(Universidad de la Salle, Colombia)

PhD. Manuel Velasco Pernía
(Universidad Central, Venezuela)

PhD. Yasmile Navarro
(Universidad del Zulia, Venezuela)

PhD. Isaac Uribe Alvarado
(Universidad de Colima, México)

PhD. Adreana Morón Vásquez
(Universidad del Zulia, Venezuela)

PhD. Francisco Mustieles
(Universidad de las Américas, UDLAP, México)

PhD. Armando J. Urdaneta Montiel
(Universidad Nacional Experimental
Rafael María Baralt, Venezuela)

PhD. Nancy Rivas de Prado
(Universidad de los Andes, Venezuela)

Dedicatoria

*A nuestras familias e hijos,
que son a quienes debemos los vestigios
de luz que orientan, dirigen e iluminan
el cauce en el camino y su transitorio
paso por él.*

Contenido

Resumen	17
Prólogo	19
Capítulo I	
Líneas de investigación en ingeniería de sistemas	23
1.1. Introducción	23
1.2. Investigación en ingeniería de sistemas	33
1.3. Sistemas de procesamiento de información	38
1.4. Sistemas inteligentes	43
1.5. Sistemas y arquitectura de procesos.....	47
1.6. Método y estandarización de sistemas.....	49
1.7. Cibernética	51
1.8. Sistemas de comunicaciones	53
Actividades del capítulo I	55
Capítulo II	
Metodología investigativa en ingeniería de sistemas	56
2.1. Fundamentos metodológicos de una investigación	56
2.2. Diseño y contextualización de una investigación	58
2.3. Metodologías investigativas cuantitativas.....	76
2.4. Metodologías investigativas cualitativas	88
2.5. Metodologías investigativas alternativas.....	90
Actividades del capítulo II.....	102

Capítulo III

Análisis estadístico aplicado a la investigación	104
3.1. Metodología estadística aplicada a la investigación	104
3.2. Fases de un análisis estadístico aplicado a la investigación ...	111
3.3. Forma de seleccionar una muestra de datos	121
3.3.1. Muestreo aleatorio simple	122
3.3.2. Muestreo aleatorio sistemático	123
3.3.3. Muestreo aleatorio estratificado	128
3.3.4. Muestreo aleatorio por conglomerado (clúster)	131
3.3.5. Muestreo no probabilístico	133
3.4. Forma de calcular el tamaño una muestra de datos	135
3.5. Estimadores poblacionales	145
3.5.1. Medidas de tendencia central	145
3.5.2. Medidas de variabilidad o dispersión	149
3.5.3. Medidas de posición (cuartiles)	153
3.5.4. Gráfico boxplot o de caja	156
3.5.5. Proporción de éxitos (\hat{P})	158
Actividades del capítulo III	161

Capítulo IV

Sistema informático biomédico inteligente en una TD	166
4.1. Tarjetas de desarrollo TD	167
4.2. Análisis la base de datos de las enfermedades hepáticas	169
4.3. Técnica de inteligencia artificial utilizada	173
4.4. Diseño de la aplicación como apoya al diagnóstico	189
4.5. Algoritmo de clasificación inteligente para el diagnóstico	192
4.6. Análisis de los resultados de las pruebas supervisadas ...	195
Referencias	199
Anexo	203

Índice de tablas

Tabla 1. Algunas formas de clasificar las ciencias	28
Tabla 2. Algunas formas de clasificar las investigaciones en ciencia y/o tecnología	30
Tabla 3. Algunas denominaciones de tipos de investigación relacionadas con proyectos o diseños	32
Tabla 4. Uso de recursos digitales en el aprendizaje de las matemáticas (1)	63
Tabla 5. Uso de recursos digitales en el aprendizaje de las matemáticas (2)	65
Tabla 6. Frecuencia de uso de recursos digitales en el aprendizaje de las matemáticas (1).....	68
Tabla 7. Frecuencia de uso de recursos digitales en el aprendizaje de las matemáticas (2).....	70
Tabla 8. Frecuencia de resultados del lanzamiento de un dado (1).....	72
Tabla 9. Frecuencia de resultados del lanzamiento de un dado (2)	73
Tabla 10. Muestra física de 27 jóvenes adolescentes con síntomas de anorexia.....	106
Tabla 11. Signos visibles de 27 jóvenes adolescentes con síntomas de anorexia.....	107
Tabla 12. Muestra aleatoria de un estrato de 45 jóvenes de sexo femenino	109

Tabla 13. Preferencias electorales de un estrato de 45 jóvenes de sexo femenino	110
Tabla 14. Comparación del nivel de destreza numérica por carrera (1).....	115
Tabla 15. Comparación del nivel de destreza numérica por carrera (2).....	117
Tabla 16. Comparación del nivel de destreza numérica por carrera (3)	117
Tabla 17. Características medibles de la población física.....	124
Tabla 18. Muestra sistémica 1.....	126
Tabla 19. Muestra sistémica 2.....	127
Tabla 20. De nivel de confianza	137
Tabla 21. Estaturas de los estudiantes de la U.S.B	140
Tabla 22. Distribución de frecuencia de las estaturas de los estudiantes de la U.S.B.....	142
Tabla 23. Límites reales del histograma y el polígono de frecuencias	144
Tabla 24. Estatura de 50 estudiantes	157
Tabla 25. Fumadores y no fumadores de la tasca.....	159
Tabla 26. Base de datos hepática	170
Tabla 27. Patrón de un paciente específico	171
Tabla 28. Valores normales de los síntomas o variables	171
Tabla 29. Toma de exámenes (Vector línea 1).....	172
Tabla 30. Toma de exámenes (Vector línea 2).....	173
Tabla 31. Matriz de confusión.....	185
Tabla 32. Tabla de requerimientos	190
Tabla 33. Datos de pruebas supervisadas SVM C++	193
Tabla 34. Datos de pruebas supervisadas “RNA” Python.....	193

Tabla 35. Datos de pruebas supervisadas “SVM” con procesamiento paralelo Python. 194

Tabla 36. Datos de pruebas supervisadas “SVM” con procesamiento en cascada Python 194

Tabla 37. Datos de pruebas realizadas por el especialista de la salud 196

Índice de figuras

Figura 1. Las cuatro dimensiones de la ingeniería.	25
Figura 2. Diseño y Contextualización de una Investigación ...	59
Figura 3. Visualización de los resultados de recursos digitales	65
Figura 4. Uso de recursos digitales en el aprendizaje de las matemáticas	66
Figura 5. Porcentaje de uso de recursos digitales en el aprendizaje de las matemáticas	66
Figura 6. Frecuencia de resultados del lanzamiento de un dado	73
Figura 7. Metodología estadística aplicada a la investigación .	105
Figura 8. Signos visibles de 27 jóvenes adolescentes con síntomas de anorexia barra	108
Figura 9. Signos visibles de 27 Jóvenes adolescentes con síntomas de anorexia torta.....	108
Figura 10. Preferencias electorales de un estrato de 45 jóvenes de sexo femenino	110
Figura 11. Estratificación de la población de estudiantes de ingeniería de la U.S.B. autores	115
Figura 12. Comparación del nivel de destreza numérica por carrera (1).....	119
Figura 13. Comparación del nivel de destreza numérica por carrera (2).....	119

Figura 14. Muestreo aleatorio sistemático (1).....	125
Figura 15. Muestreo aleatorio sistemático (2).....	127
Figura 16. Muestreo aleatorio estratificado	130
Figura 17. Ficha técnica de un muestreo estratificado.....	130
Figura 18. Muestreo aleatorio por conglomerado (clúster)...	131
Figura 19. Muestreo intencional o de conveniencia.....	134
Figura 20. Tamaño una muestra de datos	136
Figura 21. Histograma y el polígono de frecuencias.....	144
Figura 22. Ilustración de la desviación estándar en la curva normal	151
Figura 23. Cuartiles.....	154
Figura 24. Gráfico boxplot o de caja 1.....	156
Figura 25. Gráfico boxplot 2 o de caja de las estaturas en pulgadas.....	158
Figura 26. Porcentaje de fumadores y no fumadores	160
Figura 27. Tarjeta de desarrollo	168
Figura 28. Datos cargados y clasificados usando el Kernel cuadrático	176
Figura 29. Datos cargados y clasificados usando el Kernel polinómico	176
Figura 30. Datos Cargados y clasificados usando el Kernel perceptrón (MLP)	176
Figura 31. Error ϵ_i de clasificación	177
Figura 32. Comandos para obtener el modelo de entrenamiento	179
Figura 33. Esquema de una red neuronal.....	180
Figura 34. Diagrama de bloques clasificación de la red neuronal	184

Figura 35. Metodología SCRUM para desarrollo de software -
Aplicaciones complejas..... 189

Figura 36. Diagrama de casos de usos 190

Figura 37. Diagrama de secuencia..... 191

Figura 38. Muestra datos de entrenamiento y datos
de prueba 193

Figura 39. Comparación entre los resultados de las pruebas
supervisadas. 195

METODOLOGÍA DE INVESTIGACIÓN CON CONTENIDOS ESTADÍSTICOS PARA EL PROGRAMA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD SIMÓN BOLÍVAR

José Gerardo Chacón Rangel, Nahid Antuan Bautista Vega

Resumen

La presente investigación propone una metódica para el desarrollo de proyectos de ingeniería de sistemas viables, con contenidos estadísticos, aplicados al área de la salud como apoyo a la formación de investigadores en el nivel de pregrado de la Universidad Simón Bolívar sede Cúcuta. La estructura de la propuesta está organizada en cuatro capítulos. El capítulo uno presenta las Líneas de investigación en Ingeniería de Sistemas así como también algunas formas de clasificar las investigaciones en ciencia y/o tecnología según diversos autores. También se muestra los conceptos y definiciones de metodologías de investigación en ingeniería de sistemas, investigación en ingeniería de sistemas, sistemas inteligentes, de procesos y comunicaciones, método y estandarización de sistemas y cibernética. El capítulo dos muestra los fundamentos metodológicos de una investigación y la contextualización de una investigación describiendo puntualmente cada uno de los paradigmas de investigación en ingeniería de sistemas. El capítulo tres presenta el concepto de análisis estadístico aplicado a la investigación ilustrando aplicaciones prácticas del mismo y finalmente en el capítulo cuatro se presenta el desarrollo del sistema informático biomédico inteligente en una tarjeta de desarrollo aplicando técnicas básicas de ingeniería de sistemas sobre base de datos hepática siguiendo la metódica de los investigadores.

Palabras clave: Sistemas viables, cibernética, sistema informático biomédico.

Prólogo

El “ingeniero de sistemas” debe asumir su papel protagónico innovando y desarrollando nuevas herramientas tecnológicas, conducentes al desarrollo de un conocimiento aplicado que se revierta en el colectivo que hace parte de la sociedad y la nación. De ahí, la pretensión de escribir un libro para ser usado como instrumento para plantear el “*diseño y contextualización de una investigación*”, dándole un soporte al ingeniero investigador en metodología investigativa al aplicar en el diseño una estructura teórica que va a representar la “*columna vertebral de un proyecto de investigación*” debido a que ésta, le da el soporte teórico a los resultados y conclusiones obtenidas para que los mismos tengan una “*validez objetiva y científica*”, con el objetivo de dar respuesta a las interrogantes y dificultades planteadas en un problema objeto a estudio, debido a que el bosquejo y delineación esquemática estratégica de una investigación va a indicar secuencialmente los pasos a seguir para alcanzar los objetivos de la investigación y los métodos de cómo lograr dar respuesta a las incógnitas o hipótesis del problema planteado.

Para ello, en el texto se planteó primeramente la metodología investigativa en ingeniería de sistemas en donde se abordaron y explicaron las líneas investigativas inmersas tales como: los sistemas de procesamiento de información, los sistemas inteligentes, los sistemas y arquitectura de procesos, el método y estandarización de sistemas, la cibernética y los sistemas de comunicaciones. Luego, como herramientas para abordar el diseño de la investigación en alguna de las líneas investigadas expuestas, se esbozó una metodología investigativa abordándose el análisis estadístico aplicado a la investigación, el

diseño y contextualización de una investigación, las “*metodologías investigativas cuantitativas, cualitativas y holísticas*”. Una vez esbozado el diseño de la investigación, se abordaron “*estrategias y mecanismos para tomar una muestra de datos*” a ser utilizados en una investigación, exponiéndose los tipos de muestreo y el cálculo del tamaño de la muestra, para luego profundizar en la “*forma de organizar los datos de la muestra y su representación gráfica*”, teniéndose en cuenta la cantidad y forma de los datos. Después, se hizo hincapié en la importancia de los “*estimadores poblacionales*” que dan información respecto al centro de gravedad, el grado de variabilidad o dispersión y la posición relativa de los mismos en términos porcentuales.

Por último, “*se expone una investigación realizada en ingeniería de sistemas*”, en el que desarrolló un “*sistema informático*” basado en técnicas de “*inteligencia artificial*” implementado sobre una “*tarjeta de desarrollo*” para el diseño de un “*dispositivo biomédico*” para identificar y clasificar con rapidez y economía el “*diagnóstico de una enfermedad hepática*”; mediante un sistema computacional inteligente que permite diagnosticar enfermedades como la “*colecistitis hepática*” y la “*cirrosis hepática*” en tiempo real, con un mínimo de error, cuya pretensión es servir de apoyo en el “*diagnóstico de enfermedades hepáticas*” previa obtención de exámenes de química sanguínea realizados por un laboratorio clínico, aportándose así, una nueva herramienta médica para los profesionales del área de la medicina.

De ahí, que el texto represente una herramienta que apoye al desarrollo del estudiante de ingeniería de sistemas concebido como un futuro profesional íntegro y proactivo, con sentido investigativo y optimizador, capacitado para la formulación de proyectos inherentes a las áreas de gerencia, análisis de sistemas y tecnologías de la información, aplicando técnicas, estrategias y herramientas especializadas en el área capaz de aplicar los conocimientos adquiridos para:

1. La dirección de la formación de equipos multidisciplinarios, orientados a la resolución de problemas organizacionales específicos mediante la formulación de propuestas gerenciales.

2. El estudio de procesos, sistemas y situaciones que puedan ser objeto de modelado, con el fin de realizar el estudio de factibilidad técnica, económica y operativa de propuestas de solución orientadas al campo de las tecnologías de información”.
3. El análisis, diseño e implementación de sistemas de control, información, automatización o producción, integrando los departamentos, usuarios u organizaciones involucradas a los mismos.
4. El estudio de las tecnologías de información para la formulación de estudios metodológicos aplicables a la enunciación o selección de soluciones tecnológicas estables e innovadoras.
5. La estandarización de sistemas de proceso, mediante la implementación de estándares de control y documentación, tratamiento de no conformidades, y auditorías internas, para regular los patrones de productividad establecidos.
6. El estudio de los métodos de optimización de modelos matemáticos que generen patrones de comportamiento organizativo o relacional, para el estudio de sistemas complejos o no comunes, el estudio de tendencias académicas, para la formulación, planificación transmisión y evaluación de estrategias de enseñanza aprendizaje en el campo científico-tecnológico, para ser aplicadas en el campo de la docencia universitaria.
7. El manejo de recursos de información en una formación integradora, eficiente y efectiva, para maximizar su función de soporte en los procesos de toma de decisiones de la gerencia.
8. Mantener una visión global del mundo, estableciendo estándares de interdependencia entre los sistemas existentes, para la promoción de patrones de integración entre los mismos.

*José Gerardo Chacón Rangel
y Nahid Antuan Bautista Vega
Universidad Simón Bolívar Sede Cúcuta*

Líneas de investigación en ingeniería de sistemas

1.1. Introducción

En el campo de la investigación en ingeniería suelen surgir interrogantes relacionadas con lo que significa ser un ingeniero (Ara-cil, 2006) y con la creación de conocimiento en ingeniería (Larsen, 2005). Entre las diversas cuestiones que pueden plantearse, se busca encontrar respuestas dadas por diversos autores a preguntas tales como:

¿Qué relación existe entre ingeniería y ciencia?; ¿qué tipos de investigación pudieran realizarse en ingeniería?; ¿qué es un método de investigación en ingeniería orientado al diseño?; ¿cuáles técnicas de recolección y análisis de datos pueden utilizarse en un proceso de investigación orientado al diseño en ciencias de la ingeniería?; ¿cómo se evalúan (o validan) los productos de investigación de ingeniería orientados al diseño?; ¿cómo puede ayudar la estadística a la investigación de ingeniería orientada al diseño?

Para dar respuesta a estas interrogantes se expondrán comparaciones y análisis de argumentos sustentados en obras escritas por filósofos, epistemólogos, científicos e ingenieros. La primera interrogante que plantea la relación existente entre ingeniería y ciencia, conduce a plantear tres nuevas preguntas: ¿qué es *ingeniería*?, ¿qué es *ciencia*? y, ¿qué es *ciencia de la ingeniería*?

La ingeniería en su concepto ha experimentado cambios con el transcurrir de las eras. Su origen se remonta a las primeras y antiguas

civilizaciones asociándose a las labores rudimentarias de minería, la utilización del fuego, la elaboración de armas, la manufactura de herramientas para la caza y la fabricación de instrumentos que han contribuido al desarrollo de éstas civilizaciones: la rueda, la palanca y la polea, entre otros (Bucci y Terán, 2008). Sin embargo, debe señalarse que en sus albores, la ingeniería estuvo vinculada casi exclusivamente a actividades militares, gubernamentales o de orden religioso. Así, la ingeniería se encargaba de la construcción de caminos, puentes, puertos, fortalezas, templos y cementerios. En tiempos de paz, la ingeniería estaba al servicio y bienestar de la sociedad, y para el siglo XIX, algunas universidades comenzaron a ofrecer programas de estudio con el nombre de ingeniería civil en distinción clara con la ingeniería militar ejercida para la época.

Según Kirby et al. (1990), la primera definición formal de ingeniería es la presentada en 1828 por Thomas Tredgold, ingeniero y escritor inglés, quien la define como el arte de dirigir los grandes recursos de energía de la naturaleza para el uso y conveniencia del hombre, “*The art of directing the great sources of power in nature for the use and convenience of man*”. Sin embargo, muchas de las definiciones actuales de ingeniería incorporan aspectos relacionados con las ciencias, las tecnologías, las matemáticas, la innovación, la sociedad y el desarrollo sostenible, entre otras cosas; variando desde conceptos muy concisos, hasta otros muchos más amplios que asumen un carácter integral, permitiendo tener una mejor percepción del alcance de la misma.

Según el Diccionario de la Real Academia Española, DRAE (2014), *ingeniería* es el estudio y aplicación, por especialistas, de las diversas ramas de la *tecnología*, así como también la actividad profesional del ingeniero. Una acepción pertinente del término *tecnología* encontrado en el DRAE (2014) es: *Conjunto de teorías y de técnicas que permiten el aprovechamiento práctico del conocimiento científico*. Por otro lado, según el Diccionario Oxford (2014), la *ingeniería* es la rama de la *ciencia y la tecnología* dedicada al diseño, construcción, y uso de ingenios, máquinas y estructuras. De acuerdo a Marjoram y Zhong (2010),

la ingeniería es el campo o disciplina, práctica, profesión y arte que se relaciona con el desarrollo, adquisición y aplicación de conocimiento técnico, científico y matemático acerca de la comprensión, diseño, desarrollo, invención, innovación y uso de materiales, máquinas, estructuras, sistemas y procesos para propósitos específicos.

Dias (2008, 2014) señala que aunque la ingeniería es considerada hoy como diferenciada de la ciencia, la predominancia de los componentes de la ciencia básica en la educación de los ingenieros contribuye implícitamente a transmitir la idea de que la ingeniería es, en esencia, un poco más que la mera aplicación de ciencias naturales y exactas a la realidad de la práctica. Para ayudar a desafiar esta visión y contribuir a una reflexión sobre la epistemología de la ingeniería, propone un modelo donde la ingeniería es vista como un desarrollo en cuatro dimensiones transdisciplinarias: de las ciencias básicas, de las ciencias sociales, del diseño, y de la realización práctica. Esto nos permite pensar del ingeniero como un profesional que combina, en proporciones variables, las cualidades de científico, sociólogo, diseñador y hacedor, tal como se representa en la Figura 1.

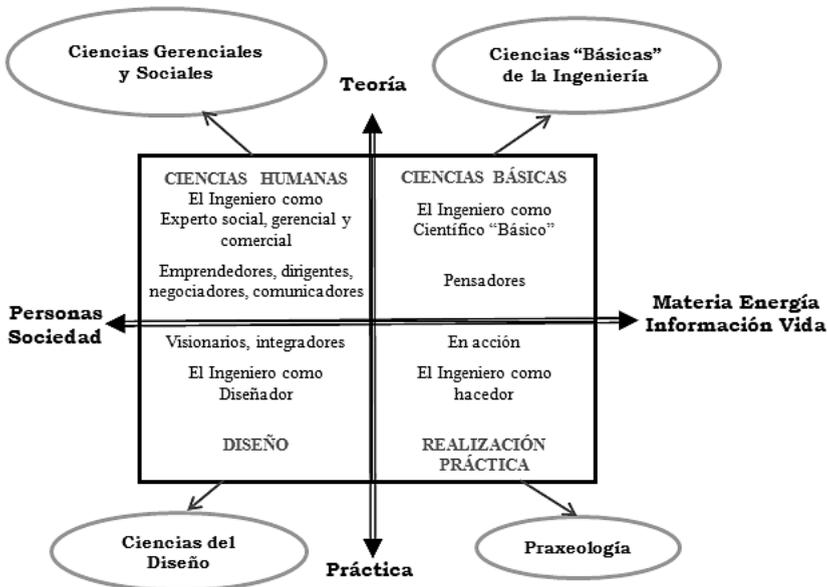


Figura 1. Las cuatro dimensiones de la ingeniería. Fuente: (Adaptado de Díaz, 2014).

La dimensión inspirada por las ciencias básicas ve la ingeniería como la aplicación de las ciencias naturales y exactas, enfatizando los valores de la lógica y el rigor, y ve el conocimiento como producido a través de análisis y experimentación. La Investigación es el *modus operandi* preferido de esta dimensión, donde el descubrimiento de los principios básicos es visto como la actividad conductora hacia un mayor reconocimiento. Con base en cuatro asuntos claves de la filosofía del conocimiento (que son los asuntos ontológicos, epistemológicos, metodológicos y axiológicos), se plantean para la ingeniería preguntas acerca de lo que *puede realmente conocer la ingeniería, qué es conocimiento de ingeniería, cómo puede construirse el conocimiento de ingeniería*, y preguntas axiológicas (que incluye la interrogante ética) sobre el valor y dignidad del conocimiento de ingeniería. Además, se enfatizan los rasgos distintivos claves que emergen de la fuerte presencia de una dimensión de diseño, que incluye la importancia ligada al razonamiento abductivo y la aceptación de cursos de acción que se aprovechan de información casual, adoptan ideas caprichosas y provocan saltos creativos que parecen ir contra el rigor científico tradicional.

Según Coyle, Murphy y Grimson (2005), el término “ciencia aplicada” ha sido a veces introducido para explicar cómo están relacionadas la ingeniería y la ciencia. El enfoque utilizado usualmente es argumentar que la ingeniería toma el conocimiento descubierto por la ciencia y lo aplica para resolver problemas en beneficio de la sociedad. Simon H. (*The Sciences of the Artificial*, 1996), indica que históricamente y tradicionalmente ha sido tarea de las disciplinas de las ciencias enseñar acerca de las cosas naturales: como son ellas y cómo funcionan, y que ha sido tarea de las escuelas de ingeniería enseñar acerca de las cosas artificiales: cómo hacer artefactos que tengan propiedades deseadas y cómo diseñarlas.

De acuerdo a Gregor (2009) las ciencias de lo artificial forman una tercera forma de ciencia, en adición a las ciencias naturales y humanas. Las ciencias de lo artificial pueden ser consideradas una ciencia debido a que usan métodos de observación de tipo científico y ex-

perimentos para probar la teoría. La teoría en la ciencia de lo artificial se relaciona con artefactos que son construidos con algún propósito, y en la teorización es necesario estudiar la relación entre los medios para alcanzar algún resultado y el resultado obtenido.

La ciencia como concepto, según el Diccionario Oxford es la actividad intelectual y práctica que abarca el estudio sistemático de la estructura y comportamiento del mundo físico y natural a través de observación y experimentación. Según Johannesson, P. y Perjons, E. (2014), la investigación científica produce diferentes tipos de conocimiento. El conocimiento puede ser utilizado para diferentes propósitos, tales como describir, explicar, y la predicción de fenómenos. Igualmente, señala que los principales tipos de conocimiento son: El conocimiento de definición (o conocimiento conceptual), que incluye conceptos, constructos, terminologías, definiciones, clasificaciones, vocabularios, y taxonomías. El conocimiento descriptivo, que describe y analiza una realidad ya existente o del pasado. El conocimiento explicativo, que proporciona respuestas a “cómo” y “por qué” y explica cómo se comportan los objetos y por qué se producen los eventos, a menudo mediante relaciones causales. El conocimiento predictivo, que predice los resultados basados en los factores subyacentes, pero sin explicar causal u otras relaciones entre ellas. El conocimiento explicativo y predictivo, que ofrece a la vez explicaciones y predicciones; El conocimiento prescriptivo, que se compone de los modelos y métodos prescriptivos, que ayuda a resolver problemas prácticos. Diversos autores han clasificado las ciencias según sus objetivos y alcance, de acuerdo al fenómeno estudiado, a la manera de producir el conocimiento científico o a su objeto o tema, tal como se muestra en la Tabla 1.

Johannesson, Perjons y Bider (2013), indican que la ciencia natural estudia los fenómenos naturales; la ciencia social estudia el comportamiento social de los humanos y las sociedades; la ciencia formal incluye matemáticas, lógica y estadísticas, las cuales utilizan pruebas en lugar de métodos empíricos como métodos de verificación; la ciencia del diseño, no solo describe y explica sino que además quiere

Tabla 1. Algunas formas de clasificar las ciencias

Referencia	Criterio de clasificación	Tipos
Jaimes (2012) Sabino (1992)	Según sus objetivos y alcance	<ul style="list-style-type: none"> • Pura • Aplicada
Jaimes (2012)		<ul style="list-style-type: none"> • Empíricas o fácticas (naturales y sociales)
Méndez (2011) Sabino (1992)	Según su objeto o tema (a)	<ul style="list-style-type: none"> • Formales
Gregor (2009)		<ul style="list-style-type: none"> • Naturales • Humanas • Prácticas, o de lo artificial
Rocha et al (2012)	Dependiendo del modo de producir conocimiento científico	<ul style="list-style-type: none"> • Ciencias formales • Ciencias explicativas • Ciencias del diseño
Johannesson et al (2013)	Según su objeto o tema (b)	<ul style="list-style-type: none"> • Natural • Social • Formal • Del diseño

Fuente: Autores.

cambiar el mundo para mejorarlo y crear nuevos mundos mediante el desarrollo de artefactos que puedan ayudar a las personas a satisfacer sus necesidades y aprovechar nuevas oportunidades.

La ciencia de la ingeniería según Love (1998), es el estudio científico de la ingeniería y las prácticas científicas por medio de la cual se produce la tecnología. La investigación de ingeniería es la investigación en temas de ingeniería que resultan en teoría de ingeniería. Boon (2009) señala que la ciencia de la ingeniería es investigación en el contexto de la tecnología. Las ciencias de la ingeniería se esfuerzan por explicar, predecir u optimizar el comportamiento de dispositivos o las propiedades de diversos materiales, sean reales o posibles.

Hendricks, Jacobsen y Pedersen (2000), sugieren que la ciencia de la ingeniería no es literalmente solamente “*ciencia aplicada*”, sino que constituye de por sí un campo específico, con sus propios métodos que producen su propio conocimiento independiente de la ciencia natural o la ciencia aplicada. Además, sugieren que hay una diferencia en las suposiciones epistemológicas y ontológicas entre la ciencia pura y la ciencia de la ingeniería, parcialmente basadas en una diferencia en valores cognitivos que gobiernan sus respectivas iniciativas. Cabe ahora entrar en la interrogante sobre la relación que existe entre ingeniería, y ciencia.

Maldonado y Gómez (2012) resaltan que la ingeniería está pasando por un proceso de “*complejificación*”, el cual consiste en (a) un cambio a través del cual la ingeniería se convierte en una ciencia, y por lo tanto deja de ser una mera praxis o profesión; (b) al convertirse en una ciencia, la ingeniería puede ser considerada como una de las ciencias de la complejidad. En realidad, la complejificación de la ingeniería es el proceso por el cual la ingeniería puede ser estudiada, alcanzada y comprendida en términos de conocimiento, y no en términos de bienes y servicios. Los sistemas complejos de ingeniería y los sistemas bio-inspirados son dos expresiones de una ingeniería compleja. Según Hendricks et al. (2000), la ciencia de la ingeniería es una disciplina científica que desde el punto de vista de la epistemología y de la filosofía de la ciencia ha sido un tanto desatendida. Cuando la ciencia de la ingeniería ha estado bajo escrutinio filosófico ha surgido la pregunta de si la ingeniería es una derivación de ciencia pura y aplicada y sus métodos. Los autores sostienen que la ingeniería es una ciencia gobernada por su propia epistemología, metodología y ontología. De esta forma, en estudios de ingeniería pudieran realizarse diversos tipos de investigación, según sea el objetivo, el uso que se le da a la información y el nivel de profundidad. La Tabla 2 mostrada a continuación, presenta diferentes tipologías de acuerdo a diversos autores:

Tabla 2. Algunas formas de clasificar las investigaciones en ciencia y/o tecnología

Referencia	Criterio de clasificación	Tipos
Hurtado (2012)	Con base en los objetivos de la investigación	<ul style="list-style-type: none"> • Exploratoria • Descriptiva • Analítica • Comparativa
Hurtado (2010)		<ul style="list-style-type: none"> • Explicativa • Predictiva • Proyectiva • Interactiva • Confirmatoria • Evaluativa
Turrioni y Mello (2012)	Según los objetivos	<ul style="list-style-type: none"> • Exploratoria • Descriptiva • Explicativa • Normativa
Phillips y Pugh (2008)		<ul style="list-style-type: none"> • Exploratoria o preparatoria • De comprobación • De solución de problemas
García Córdoba (2007)	De acuerdo con el uso de la información	<ul style="list-style-type: none"> • Pura • Aplicada • Tecnológica • Desarrollo

Tabla 2 (Continuación)

Referencia	Criterio de clasificación	Tipos
Turrioni y Mello (2012)	Según su naturaleza	<ul style="list-style-type: none"> • Básica • Aplicada
Turrioni y Mello (2012)	Según su objetivo	<ul style="list-style-type: none"> • Exploratoria • Descriptiva • Explicativa • Normativa
Turrioni y Mello (2012)	Según el abordaje	<ul style="list-style-type: none"> • Cuantitativa • Cualitativa • Combinada
Balestrini (2002)	Según el tipo de estudio	<ul style="list-style-type: none"> • Formulativo o exploratorio • Descriptivo • Diagnóstico • Explicativo • Proyecto Factible

Fuente: Autores.

Los tipos de investigación que se relacionan más con diseños o proyectos, se ilustran en la Tabla 3.

Tabla 3. Algunas denominaciones de tipos de investigación relacionadas con proyectos o diseños

Denominación del tipo de investigación	Referencia
Investigación tecnológica	<ul style="list-style-type: none"> • García Córdoba (2007) • García Córdoba et al. • Ramírez (2010)
Solución de problemas	<ul style="list-style-type: none"> • Phillips y Pugh (2008)
Investigación normativa	<ul style="list-style-type: none"> • Turrioni y Mello (2012)
Proyecto factible	<ul style="list-style-type: none"> • UPEL (2011) y Balestrini (2002)
Investigación proyectiva	<ul style="list-style-type: none"> • Hurtado (2012) • Hurtado (2010)
Investigación orientada al diseño, (Design-oriented Research), o Investigación en ciencias del diseño (Design Science Research)	<ul style="list-style-type: none"> • Verschuren (2005) • Peffers, Tuunanen, Rothenberger & Chatterjee (2008) • Johannesson y Perjons (2014) • Dresch, Lacerda y Antunes (2015) • Vaishnavi y Kuechler (2015)

Fuente: Autores.

La investigación de carácter tecnológico es un proceso cuya finalidad es proporcionar un “*saber hacer*” con la realidad para transformarla y obtener beneficios; es una labor que aun cuando genera conciencia de la complejidad de la realidad, potencia la confianza del “*saber cómo*”, apoyado fundamentalmente en la ciencia y la tecnología (García Córdoba, 2007). El proceso de la investigación tecnológica no se centra en explicar una realidad concreta, sino que trata de ir de las ideas a las acciones para generar bienes o servicios que facilitan la vida. Persigue un conocimiento práctico: un conjunto de instrucciones a seguir para transformar el objeto. “Los resultados de una investigación tecnológica son conocimientos que establecen con detalle: el qué, el cómo y con qué lograr los objetivos prede-

terminados” (García Córdoba, F.; Morales, A.; Muñoz, C. 2010). Las investigaciones tecnológicas suponen la elaboración de prototipos o diseños para ser implementados y desarrollados con el fin de resolver problemas reales o una situación determinada. Estos estudios deben incluir la factibilidad de aplicación del prototipo o diseño (Ramírez, 2010). En el tipo de investigación denominado investigación de solución de problemas se parte de un problema concreto en el mundo real y se reúne todos los recursos intelectuales que pueden influir en su solución. El problema tiene que ser definido y el método de solución ha de ser descubierto (Phillips y Pugh, 2008).

1.2. Investigación en ingeniería de sistemas

El “ingeniero de sistemas” debe asumir un papel protagónico en la sociedad en donde se desenvuelve, innovando y desarrollando tecnologías aplicadas de vanguardia, conducentes al desarrollo de conocimiento y de la sociedad en función de las necesidades del país, dado que las competencias industriales en cualquier país del mundo están determinadas por el avance tecnológico, es decir, la capacidad de los países en adquirir, asimilar, adecuar e innovar la ciencia y la tecnología aplicadas a sus propios procesos de producción y servicios determinado así su riqueza y fortaleza.

Es por ello, que el estudiante de “ingeniería de sistemas” debe ser auspiciado a través de su inmersión en la investigación, en el desarrollo de su capacidad de razonamiento abstracto y espacial con manejo de las tecnologías analógicas y digitales de vanguardia, capacitado para captar relaciones espaciales, en virtud de su formación en ciencias abstractas para interpretar y aplicar estos principios a situaciones prácticas, propiciando su desarrollo en “razonamiento lógico e imaginación creadora”, con una alta precisión y minuciosidad a la hora de resolver problemas prácticos y aplicados derivados de procesos abstracción, con capacidad plena de desempeñarse como un “investigador” que persiga descubrir nuevos conocimientos en el ámbito científico y tecnológico, como un “desarrollador” que aplique los resultados de la investigación para la fabricación de nuevos

productos tecnológicos para el desarrollo social y de carácter “innovador”, por lo que el resultado de su investigación representa un avance e innovación tecnológica que mejore los productos tecnológicos existentes.

Lo planteado anteriormente reviste la importancia y la necesidad de un profesional que maneje los elementos que contempla “el proceso de investigación y su metodología” que le servirán de vehículo para la toma de decisiones. Lo anterior ha de afianzar su perfil entendido como “un profesional íntegro, proactivo, con sentido investigativo y optimizador”, “capacitado para la formulación de proyectos inherentes a las áreas de gerencia, análisis de sistemas y tecnologías de la información, aplicando técnicas, estrategias y herramientas en la dirección de equipos multidisciplinarios, orientados a la resolución de problemas organizacionales específicos mediante la formulación de propuestas gerenciales”, “capacitado para el estudio de procesos, sistemas y situaciones que puedan ser objeto de modelado, con el fin de realizar el estudio de factibilidad técnica, económica y operativa de propuestas de solución orientadas al campo de las tecnologías de información”, “capacitado para el análisis, diseño e implementación de sistemas de control, información, automatización o producción, integrando los departamentos, usuarios u organizaciones involucradas a los mismos”, “capacitado para el estudio de las tecnologías de Información para la formulación de estudios metodológicos aplicables a la enunciación o selección de soluciones tecnológicas estables e innovadoras”, “capacitado para la estandarización de sistemas de proceso, mediante la implementación de estándares de control y documentación, tratamiento de no conformidades, y auditorías internas, para regular los patrones de productividad establecidos”, “capacitado para el estudio de los métodos de optimización de modelos matemáticos que generen patrones de comportamiento organizativo o relacional, para el estudio de sistemas complejos o no comunes”, “capacitado para el estudio de tendencias académicas, para la formulación, planificación transmisión y evaluación de estrategias de enseñanza aprendizaje en el campo científico-tecnológico, para ser aplicadas en el campo de la

docencia universitaria”, “capacitado para el manejo de recursos de información en una formación integradora, eficiente y efectiva, para maximizar su función de soporte en los procesos de toma de decisiones de la gerencia” y en “capacitado para mantener una visión global del mundo, estableciendo estándares de interdependencia entre los sistemas existentes, para la promoción de patrones de integración entre los mismos”.

Lo anteriormente expuesto, asegura su exitoso desempeño en empresas, compañías e industrias del sector público o privado, lo que requerirá implementar sistematización, automatización, control y optimización de procesos a gran escala, dado que la aplicación de “teoría de sistemas” está presente en todas las actividades humanas en la mejora de la efectividad de procesos humanos y empresariales. Es importante mencionar que uno de los grandes retos del “ingeniero de sistemas” es la “innovación”, y esto pasa por un proceso previo de “investigación” a la hora de crear de soluciones nuevas, en un “enfoque interdisciplinario” que permita estudiar y comprender la realidad, con el propósito de implementar u optimizar sistemas complejos, incorporar mecanismos para optimizar al evaluar costo, evaluar la efectividad de recursos a implementarse en un sistemas, formular planes para integrar diferentes etapas de un proceso de desarrollo, gerenciar a través del diseño, análisis y control de un sistema; transformando una necesidad de operación en una descripción de parámetros de rendimiento del sistema y una configuración del sistema a través del uso de un proceso iterativo de definición, síntesis, análisis, diseño, prueba y evaluación; integrando parámetros técnicos relacionados para asegurar la compatibilidad de todas las interfaces de programa y funcionales de manera que optimice la definición y diseño del sistema total; integrando factores de fiabilidad, mantenibilidad, seguridad, supervivencia, humanos y otros en el esfuerzo de ingeniería total a fin de cumplir los objetivos de coste, planificación y rendimiento técnico. De ahí, un hecho relevante y de vital importancia, que futuro “ingeniero de sistemas” adquiera conocimientos de “metodología de la investigación” para el planteamiento y ejecución

de una investigación en pro de un “*paradigma sistémico del desarrollo humano de la ciencia y de la tecnología*” con capacidad de adaptación a la multiplicidad de situaciones investigativas planteadas en un mundo globalizado e intercomunicado, integrando otras disciplinas y grupos de especialidad en un esfuerzo de equipo, formando un proceso de desarrollo estructurado.

La importancia de la Investigación y su metodología, es que representa una de las herramientas más efectivas para alcanzar la competitividad y la excelencia académica, permitiéndose el desarrollo de habilidades, destrezas y actitudes que se requieren para construir datos, información y conocimiento, ya que constituye una actividad creadora y productora de nuevas ideas, debido a que en la actualidad la ciencia y la tecnología se han abanderado como pilares para el desarrollo social, y el manejo ya no de la información sino del conocimiento, hace necesario estar capacitados para generarlo constantemente y retenerlo en beneficio de la comunidad, incorporando el desarrollo tecnológico en dónde fluyen metodologías, modelos y teorías que se deben saber organizar y utilizar.

El liderazgo en ciencia y tecnología determinan el nuevo orden económico mundial. Las competencias industriales en cualquier país del mundo están determinadas por el avance tecnológico, es decir que la capacidad que tengan de adquirir, asimilar, adecuar e innovar la ciencia y la tecnología aplicadas a sus propios procesos de producción y servicios determinan la riqueza y fortaleza de los países. La mejor manera de afrontar el proceso natural evolutivo es la creación de conocimiento y este se logra mediante una constante investigación por ser el principal mecanismo para la solución de un problema y constituye un estímulo para la actividad intelectual creadora. También se debe mencionar que una de las ventajas de los “*ingenieros de sistemas*” es no solo poseer las características de personas creativas, sino también poseer el conocimiento técnico para que una idea sea funcional y el pensamiento lógico para llevarla a cabo, por ser un individuo capaz de combinar la inteligencia, el conocimiento y la creatividad para generar soluciones, muchas de ellas innovadoras.

De ahí la relevancia el auspicio, a través de la investigación, del desarrollo de capacidades con manejo de las tecnologías “analógicas y digitales” de vanguardia, para interpretar y aplicar estos principios, con un gran sentido de razonamiento lógico e imaginación creadora, con una alta precisión y minuciosidad a la hora de resolver problemas prácticos y aplicados derivados de procesos abstracción.

La investigación en ingeniería de sistemas cobra relevancia con la aparición de los primeros ordenadores y sus iniciales aplicaciones en “sistemas operativos”, “hojas electrónicas y/o de cálculo” y “bases de datos”. Más tarde los esfuerzos se enfocaron hacia la industria, la automatización de empresas, la generación de modelos, las metodologías de diseño y desarrollo, la programación y el diseño orientado a objetos, la lógica difusa, la teoría de fractales y los algoritmos genéticos. Además de las áreas directamente relacionadas con: “tecnología de la información”, “toma de decisiones”, “telecomunicaciones”, “administración de recursos empresariales”, “organización y métodos”, “consultoría y auditoría de sistemas de información”, ubicándose estas en apartados como: gerencia, desarrollo de software, análisis y diseño de sistemas de información, optimización de modelos matemáticos, desarrollo abstracto y estudio de componentes estructurales. Más adelante se crean líneas de investigación en: “ingeniería de software”, “algoritmos genéticos”, “algoritmos neuronales”, “algoritmos de lógica difusa”, “desarrollo de lenguajes de programación”, “compiladores”, “estructuras de datos avanzados”, “desarrollo en comunicación de datos y sistemas operativos distribuidos y de tiempo real”, “modelamiento y simulación de sistemas utilizando dinámica de sistemas”, “planificación estratégica y desarrollo de indicadores”, “evaluación de desempeño”, “programación y diseño orientado a objetos”, “arquitectura de computadores y dispositivos de telecomunicaciones”, “informática educativa”, “computación ubicua”, “creación y administración de bases de datos”, “minería de datos”, “programación colaborativa”, “bases de conocimientos”, “inteligencia artificial”, “sistemas expertos”, “sistemas inteligentes”, “aplicaciones de la lógica difusa a problemas de optimización”, “ordenadores borrosos ex-

perimentales”, “tratamiento y representación de la incertidumbre en sistemas inteligentes”, “medición y compararon diferentes algoritmos de aprendizaje sobre problemas complejos”, “redes neuronales y lógica difusa”, “gestión de bases de datos difusas”, “identificación de sistemas mediante técnicas de aprendizaje inductivo”, “aplicaciones de la lógica difusa con algoritmos genéticos”, “modelos difusos de redes neuronales y sus aplicaciones”, “aprendizaje en entornos imprecisos e inciertos”, “automatización de los procesos organizacionales”. De acuerdo al “Manual de Líneas de Investigación de Ingeniería de Sistemas” del Departamento de Investigación y Postgrado del Politécnico “Santiago Mariño” Extensión Valencia (2012), actualmente “la investigación en ingeniería de sistemas” se diversificado actualmente en diversas líneas: *sistemas de procesamiento de información, sistemas inteligentes, sistemas y arquitectura de procesos, métodos y estandarización de sistemas, cibernética y sistemas de comunicaciones.*

1.3. Sistemas de procesamiento de información

Esta línea de investigación se orienta al estudio de aspectos formales de ingeniería de software, bases de datos, recuperación de información y desarrollo dirigido por modelos. Esto incluye aspectos teóricos y de implementación relacionados con el modelado, representación, consulta, manipulación y recuperación de datos o información desde repositorios estructurados, semi-estructurados o no estructurados. Estos repositorios abarcan bases de datos tradicionales, bases de datos espacio-temporales, bases de datos multimediales, y la Web, entre otros. Dependiendo del tipo de sistema de información que se esté tratando, las funciones esenciales que respaldan su existencia se verán modificadas. En general, los sistemas de procesamiento de información tienen como objetivo “respaldar las operaciones empresariales”, “respaldar la toma de decisiones gerenciales”, “respaldar la ventaja competitiva estratégica”, “contribuir a la automatización de actividades y procesos en las empresas” y “llevar la información de manera oportuna y adecuada a las instancias de la empresa que así lo requieran”.

Los sistemas de procesamiento de información actuales buscan atesorar piezas de información tales como libros, revistas, entre otros, para facilitar la producción de conocimiento sin la necesidad de experimentar cada parte de la realidad sobre la que se cimienta la idea. Así vemos como bibliotecas y centros de documentación, entre otras fuentes de información, facilitan el desarrollo de nuevas ideas y conceptos “conocimientos”. Los intereses principales para esta línea de investigación son:

I) Los sistemas de información espacio-temporal, que abarca el modelado, representación y manipulación de información que tiene una dimensión espacial. Diversas aplicaciones requieren el manejo de información espacial, como por ejemplo, sistemas de información geográficos, procesamiento digital de imágenes, navegación de robots, y modelado gráfico. Intereses principales para esta línea de investigación son: **i) base de datos espaciales**: involucra consistencia de datos, lenguajes de consulta espacial, procesamiento de consulta, extracción de datos, recuperación e integración de información espacial, **ii) recuperación de información espacial**: involucra aspectos de geo referenciación, indexación y procesamiento de consultas a documentos de texto o gráfico con contenido espacial y **iii) modelado espacial**: considera modelos formales del espacio y tiempo e inferencia de relaciones espaciales.

II) La hipermedia adaptativa, que busca respuestas a las dificultades que origina la gran cantidad de información que se provee a los usuarios, a través de sistemas hipermediales tradicionales y, especialmente, de aplicaciones Web. Para ello, promueve el desarrollo de sistemas que adaptan automáticamente su navegación y presentación a las características de cada usuario y del contexto de interacción, para facilitar el acceso a la información y servicios que le son más relevantes. Distintos enfoques de modelado de usuarios, modelos conceptuales de interfaces adaptativas, algoritmos de recomendación de información a los usuarios, factores espacio-temporales que condicionan el proceso de adaptabilidad, son algunos de los tópicos de ésta área.

III) Los sistemas transaccionales, que son sistemas capaces de enmendar cualquier error ocurrido durante una transacción, pudiendo deshacer las operaciones realizadas, manteniendo los datos tal cual estaban antes del error. También debe ser capaz de controlar y administrar múltiples transacciones, determinando prioridades entre éstas. Como ilustración se puede plantear lo siguiente: si un cliente está haciendo la reserva de un asiento en un vuelo, dicho asiento debe ser bloqueado temporalmente hasta que se concrete la transacción, porque otro cliente podría estar queriendo reservar el mismo asiento en el mismo momento. Debe controlar las transacciones para mantener la seguridad y consistencia de los datos involucrados. Otra ilustración: un cliente transfiere dinero de una cuenta a otra cuenta dentro de un mismo banco; la cantidad de dinero que se descuenta de la cuenta emisora debe ser igual a la que se suma en la cuenta receptora, de no ser así, la acción “transacción” no se realiza.

IV) La inteligencia de negocios, que es el conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa. Este conjunto de estrategias y herramientas tienen en común las siguientes características: accesibilidad a la información, apoyo en la toma de decisiones, orientación al usuario final. Algunas de las técnicas para aplicar la inteligencia de negocios son: *data Warehousing*, *DataMining*, *OLAP* y *Dashboard*.

V) La planificación de recursos empresariales, que está orientada hacia la automatización de las diferentes áreas operativas dentro de un entorno empresarial. Bajo el apoyo de la TI, los sistemas de información ERP (*Enterprise Resource Planning*), permiten una completa gestión financiera, analítica, productiva y de negocios, que unifica información y procesos en toda la organización. Se diversifica en: **i) finanzas**, a través de la implementación de sistemas en la automatización de la contabilidad empresarial a través de sistemas gerenciales que recogen y procesan el registro, resumen y gestión de todos los datos con trascendencia contable, centralizándolos para su consulta, publicación y control. De igual forma, se enfoca en las

finanzas empresariales, específicamente en los estados financieros para análisis y otros reportes, que permiten estimar la rentabilidad de la organización; **ii) operaciones**, a través de la implementación de sistemas que se encargan de gestionar la planificación y control de la producción, gestionar las órdenes de producción y fabricación, controlar el consumo de materias primas y recursos de fabricación, calcular las necesidades de producción según las unidades a producir, calcular y administrar las listas de materiales y mezclas, elaborar el plan maestro de producción, creación de órdenes de compras según necesidades y controlar la carga de trabajo según las capacidades de los centros de fabricación; **iii) ventas**, a través la implementación de sistemas que permiten gestionar la creación de clientes, facturación y cobros. Sus principales funciones están en la gestión de clientes y prospectos, ofertas, pedidos, facturación, listas de precios, gestión de precios internacionales, envíos y devoluciones. Gestión e implementación del comercio electrónico en sus dos modalidades B2B y B2C; Recursos Humanos, a través de la implementación de sistemas que permiten gestionar todo el proceso de planificación y cálculo de la nómina, altas y bajas, contrataciones, horarios y archivos de personal. Cálculo y gestión de beneficios de ley y otros beneficios ofertados por la organización. Además gestionar la planificación de cursos y otras actividades de mejoramiento profesional.

VI) La gestión de relaciones con los clientes (CRM), que está dirigida a la automatización de estrategias de ventas y el marketing empresarial. La gestión de relaciones con los clientes o CRM (Customer relationship management) es un conjunto de estrategias, procesos de negocios y políticas que están diseñadas para captar, retener y dar servicio a los clientes. Esto incluye todos los procesos de negocios que afectan directamente al cliente como lo son la mercadotecnia, las ventas y el servicio postventa. De igual forma, un modelo de negocios basado en CRM es un enfoque global que unifica las labores y estrategias de planificación, ventas, mercadotecnia y el servicio al cliente, integrados bajo la tecnología de la información, que permitirán establecer, mantener y ampliar las relaciones comerciales

con los clientes. Se diversifica en: **i) mercadeo – marketing**, a través de la incorporación de sistemas que automatizan todos los procesos relacionados a la promoción y mercadotecnia de bienes y servicios dentro de la empresa, cubriendo los puntos de contacto con los clientes o las estrategias relacionales utilizadas. De igual forma permiten administrar y hacer seguimiento a las labores de marketing “campañas, promociones y canales”, alimentándose de diferentes bases de datos, además de sus recursos propios de almacenamiento. Finalmente, permiten medir los resultados y la efectividad de las campañas de marketing lanzadas; **ii) servicios**, a través de la incorporación de sistemas de información CRM enfocados a servicios, que tratan de automatizar todos los aspectos relacionados al servicio de ventas y post venta. Esto incluye tratar de dar soluciones de forma automática a las dudas o problemas que puedan presentar los clientes y prospectos dentro del proceso de ventas y post ventas. Estos sistemas pueden gestionar el registro de información de productos y servicios (garantías, desperfectos y otros problemas), soluciones a problemas (*lugares donde canalizar la garantía, reparación y cambios*), toda la información necesaria para realizar compras (*características del producto o servicio, garantías, despacho, formas de pago*); **iii) analíticas**, a través de la incorporación de sistemas de información que analizan grandes volúmenes de datos con el fin de crear perfiles de clientes y otros elementos de información necesarios que faciliten la planificación, toma de decisiones y las acciones de marketing y ventas. Los referidos sistemas se apoyan en diferentes recursos como el uso de técnicas avanzadas de análisis y tratamiento de datos (*minería de datos, Olaps, algoritmos inteligentes*) así como de almacenamiento (*almacenes de datos y otras fuentes externas*) que permitan generar la información necesaria para determinar que clientes y prospectos serán más productivos para la empresa (*segmentación y personalización de clientes y mercados*); **iv) sistemas social CRM / CRM 2.0**, a través de la incorporación de sistemas que permiten gestionar relaciones con clientes a través de la web 2.0 y sus diferentes medios, buscando consolidarse como un modelo colaborativo de CRM. Don-

de cada medio web 2.0 representa un canal de comunicación con el cliente. Aquí se observan todos los puntos de contacto y procesos de negocios con el cliente: “*marketing, ventas y servicios*”. Estos sistemas buscan captar y analizar conversaciones, gustos, preferencias y otros patrones que puedan dar perfiles y conductas de compras de clientes y prospectos; promover productos, servicios y beneficios; vender dichos productos y servicios; así como captar la atención del cliente para que de forma a los nuevos cambios y diseños futuros para productos; mantener la comunicación y colaboración continua entre empresa, cliente y proveedores y cualquier otro elemento o entidad presente en la cadena de comercialización.

1.4. Sistemas inteligentes

Esta línea de investigación comprende el estudio, propuesta y experimentación de procesos adaptativos para la obtención de “*sistemas inteligentes*” capaces de funcionar adecuadamente en entornos de información sumamente cambiantes. Incluye la aplicación de diversas arquitecturas de “*redes neuronales*” a la solución de problemas de agrupamiento –*clustering*– y el estudio de arquitecturas competitivas dinámicas para la recuperación de información en sistemas distribuidos de gran escala en forma eficiente. Un sistema inteligente completo incluye “*sentidos*” que le permiten recibir información de su entorno, puede actuar, y tiene una memoria para archivar el resultado de sus acciones. Tiene un objetivo e inspeccionando su memoria puede aprender a cómo mejorar su rendimiento y eficiencia. Su objetivo es construir un artefacto (*robot, máquina, proceso informático, etc.*) que pueda representar su propio conocimiento y razonar sobre él, que pueda planificar y actuar, que pueda asimilar nuevo conocimiento de la experiencia y de la interacción con el entorno y que, en definitiva, pueda llevar a cabo cualquier tarea que consideramos como propias de los seres inteligentes, para aportar soluciones a problemas, como si de humanos se tratara, es decir capaz de mostrar soluciones inteligentes.

Esto es posible gracias a que al sistema lo crean con expertos (humanos), que intentan estructurar y formalizar conocimientos poniéndolos a disposición del sistema, para que este pueda resolver una función dentro del ámbito del problema, de igual forma que lo hubiera hecho un experto. Los avances tecnológicos en el área de los *sistemas inteligentes*, tienen ahora soportes o asideros legales que promueven su aplicabilidad, tal es el caso de nuevas disposiciones gubernamentales referentes a la incorporación de personas con discapacidades al ámbito laboral y educativo, así como la adaptación de dispositivos que les peritan realizar labores de movilización, bien sea por sus propias fuerzas o en vehículos adaptados para tales fines.

También en el ámbito comunicacional se han visto varios requerimientos legales que exigen a las empresas del ramo a contar con dispositivos y/o sistemas que faciliten el acceso a todas las personas aun cuando tengan impedimentos físicos que se los dificulte. En el ramo de la salud tanto pública como ocupacional, *“es cada vez más necesario contar con el apoyo de tecnologías y sistemas que faciliten las labores tanto de diagnóstico como de tratamiento a fin de dar soporte legal tanto a actuaciones médicas como a las funciones de los entes aseguradores”*. Finalmente en lo que corresponde a la seguridad tanto personal, como pública se tienen requerimientos cada vez mayores para el uso de técnicas biométricas que faciliten tanto la identificación de personas como la detección y prevención de posibles riesgos. Los intereses principales para esta línea de investigación son:

1) El procesamiento del lenguaje natural, que es el conjunto de parámetros, sistemas o procesos mediante los cuales se puede producir una comunicación entre una computadora y un usuario por medio del uso del lenguaje natural. Contempla técnicas estadísticas de procesamiento de textos y voz, para la diferenciación de procesos de ambigüedad léxica y etiquetado sintáctico.

Esta línea de investigación tiene diversos usos, entre los cuales se cuentan la traducción automática, resumen de textos y documentos, recuperación de archivos y componentes lingüísticos.

II) La inteligencia artificial, que se dedica a través de la computación al desarrollo de agentes racionales no vivos. La idea de construir una máquina que pueda ejecutar tareas percibidas como requerimientos de inteligencia humana es un atractivo. Las tareas que han sido estudiadas desde este punto de vista incluyen juegos, traducción de idiomas, comprensión de idiomas, diagnóstico de fallas, robótica, suministro de asesoría experta en diversos temas. La inteligencia artificial trata de conseguir que los ordenadores simulen en cierta manera la inteligencia humana. Se acude a sus técnicas cuando es necesario incorporar en un sistema informático, conocimiento o características propias del ser humano. Por ejemplo: Podemos interrogar a algunas bases de datos de Internet en lenguaje natural, o incluso charlar con ellas nuestro idioma, porque por detrás se está ejecutando un programa de inteligencia artificial. La inteligencia artificial incluye varios campos de desarrollo tales como:

i) Algoritmos genéticos, que son un logro más de la inteligencia artificial en su intento de replicar comportamientos biológicos, con los avances científicos que ello implica, mediante la computación. Se trata de algoritmos de búsqueda basados en la mecánica de la selección natural y de la genética de los organismos vivos utilizando la información histórica para encontrar nuevos puntos de búsqueda de una solución óptima del problema planteado, con esperanzas de mejorar los resultados.

ii) Lógica difusa, que a través de la inteligencia artificial proporciona una metodología simple y elegante de obtener una conclusión a partir de información de entrada vaga, ambigua, imprecisa, con ruido o incompleta, característico de muchos sistemas expertos. La lógica difusa se adapta mejor al mundo real en el que vivimos, e incluso puede comprender y funcionar con nuestras expresiones, del tipo “hace mucho calor”, “no es muy alto”, “el ritmo del corazón está un poco acelerado”.

iii) Realidad virtual, que es un sistema tecnológico de simulación de la realidad por computadora, de forma dinámica y tridimensional, con alto contenido gráfico, acústico y táctil, orientada a

la visualización de situaciones y variables complejas, durante la cual el usuario ingresa, a través del uso de sofisticados dispositivos de entrada, a “mundos” que aparentan ser reales, resultando inmerso en ambientes altamente participativos, de origen artificial.

III) Los agentes inteligentes, que son sistemas que poseen la capacidad de percibir impulsos de su entorno, procesar de manera racional dichas percepciones y reaccionar de forma inteligente ante lo que le rodea, optimizando la capacidad de respuesta efectiva. La mayoría de los sistemas inteligentes tiene la capacidad de autonomía de decisión y abstracción, lo cual los define como “*agentes inteligentes de abstracción*” o AIA. Estos normalmente se definen como sistemas capaces de aprender métodos de resolución de problemas, adaptables al medio, por lo cual se relacionan de manera cercana a los conceptos de inteligencia artificial y sistemas expertos.

IV) Los sistemas expertos, que son un derivado de la rama de la inteligencia artificial es un sistema informático diseñado para resolver problemas de un área específica, y al que de algún modo se le ha dotado de una competencia similar a la de un experto humano de esa área. La idea no es sustituir a los humanos expertos, sino que sirvan de apoyo a los especialistas en un “dominio” de aplicación específico. Estos sistemas imitan las actividades de un humano para resolver problemas de distinta índole (no necesariamente tiene que ser de inteligencia artificial).

También los sistemas expertos se basan en el conocimiento declarativo (hechos sobre objetos, situaciones) y el conocimiento de control (información sobre el seguimiento de una acción). Se diversifican en: **i) redes neuronales**, que es un nuevo sistema para el tratamiento de la información, cuya unidad básica de procesamiento está inspirada en la célula fundamental del sistema nervioso humano: la neurona. El objetivo es conseguir que las máquinas den respuestas similares a las que es capaz de dar el cerebro que se caracterizan por su generalización y su robustez. Las redes neuronales son utilizadas para la predicción, la minería de datos “*data mining*”, el reconoci-

miento de patrones y los sistemas de control adaptativo. Constituyen una parte muy importante en el estudio y desarrollo de la inteligencia artificial “AI” y el de la vida artificial “a-life”; **ii) sistema de aprendizaje**, o aprendizaje de máquinas que es una rama de la inteligencia artificial cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender. De forma más concreta, se trata de crear programas capaces de generalizar comportamientos a partir de una información no estructurada suministrada en forma de ejemplos. Es, por lo tanto, un proceso de inducción del conocimiento.

En los “sistemas de aprendizaje” se emplea una combinación de software y equipos que le permite a la computadora cambiar su modo de funcionar o reaccionar a situaciones, basado en la retroalimentación que recibe. Por ejemplo, algunos juegos computarizados tienen capacidades de aprendizaje, si la computadora no gana un juego en particular, recuerda no hacer los mismos movimientos; **iii) sistemas de monitoreo**, que se aplica para estudios que requieren de un caso particular de la interpretación, y consiste en la comparación continua de los valores de las señales o datos de entrada y unos valores que actúan como criterios de normalidad o estándares. Es decir, los sistemas de monitoreo comparan observaciones del comportamiento del sistema con el comportamiento estándar, dicho de otra forma, se compara lo actual con lo esperado y **iv) base de conocimiento**, que es un tipo especial de base de datos para la gestión del conocimiento. Provee los medios para la recolección, organización y recuperación computarizada de conocimiento.

1.5. Sistemas y arquitectura de procesos

Esta línea de investigación se emplaza hacia la generación de la identificación de procedimientos para la definición de la estructura funcional y no funcional de un proceso bajo conceptos de avanzada, que involucran la Ingeniería de software, la Ingeniería de procesos, la deducción de requisitos y el modelado de procesos sistémicos. El desarrollo bajo esta línea, permite optimizar la relación recurso-sistema-desarrollador, así como la identificación de las tecnologías a utili-

zar y el alcance real de los nuevos procesos a incorporar, detallando las etapas de identificación, análisis y desarrollo del mejoramiento continuo de los procesos.

Los cambios de paradigma de las organizaciones en el desarrollo de sus sistemas de información: de los datos a los procesos hace enfatizar los procesos de negocio para conseguir arquitecturas más ágiles y flexibles, adaptables a los continuos cambios que se producen en los mercados en los que las organizaciones desarrollan su negocio. El objetivo es independizar la gestión de los procesos de negocio de las aplicaciones, para que cualquier modificación en la lógica de negocio no afecte al código de las aplicaciones. Los intereses principales para esta línea de investigación son: **I) La reingeniería**, la cual contempla el rediseño de sistemas para lograr el máximo aprovechamiento de sus recursos, incorporando técnicas, herramientas, procesos e instrumentos innovadores con el fin de acoplar mediante estrategias de análisis objetivo, los sistemas existentes y las nuevas tecnologías de información. La reingeniería persigue el reordenamiento de los conocimientos aplicados a un proceso y no a una función, a fin de utilizar los diversos enfoques de desarrollo de sistemas de manera estructural y efectiva. Se diversifican en: **i) reingeniería de procesos**, que se refiere a la reinención de la estructura de un proceso, departamento u organización, manteniendo las finalidades corporativas de objetivos, misión, visión y estrategias de negocio en cuenta; **ii) reingeniería de software**, que está dirigida a la actualización, reformulación, conversión y modularización de un programa informático o diversos componentes del mismo, a partir del proceso de análisis jerárquico, utilizando la reducción de requisitos y las técnicas de modelado de información como base fundamental en el rediseño del componente; y la **iii) reingeniería de datos**, que reúne los procesos mediante los cuales se obtiene una base de datos a partir del modelo conceptual de una data base existente. Normalmente implica el uso de ingeniería inversa para el logro de estructuras físicas de la base de datos, para lograr el aprovechamiento de la información a utilizar en un proceso sistémico.

1.6. Método y estandarización de sistemas

Esta línea de investigación está orientada hacia la generación de herramientas desarrolladas bajo estándares informáticos de punta, que permitan la normalización y evaluación de las fases, actividades, tareas y eventos que componen los diferentes modelos de “*sistemas de Información*”. A fin de lograr este objetivo, se profundiza sobre los conceptos de normas de estandarización, tecnologías de información, evaluación de sistemas y metodologías para el desarrollo de sistemas de información, a fin de facilitar el análisis funcional de las necesidades existentes, con un enfoque holístico, conceptualizado, efectivo y confiable.

Esta línea de investigación tiene como objetivo diseñar sistemas estratégicos que permitan la evaluación del cumplimiento de los planes, programas, políticas, normas y lineamientos que regulan la actuación de los empleados y funcionarios de una institución, junto con sus sistemas de manejo de información, así como evaluar las actividades que se desarrollan en sus áreas y unidades administrativas. La estandarización y normalización de procesos, es pilar fundamental dentro de un proceso organizativo. Los intereses principales para esta línea de investigación son:

I) **La metodologías de sistemas**, que está dirigida hacia el estudio y evaluación de las metodologías de sistemas existentes, mediante un análisis crítico y objetivo, a fin de desarrollar un concepto metodológico estable, adaptativo y eficaz. Se busca un estudio cónsono con los avances dentro del marco de las TI, que contemplen etapas, estrategias, herramientas y métodos para el modelado de sistemas de información.

II) **La normalización**, que persigue la regulación de procesos sistémicos mediante el desarrollo e implementación de estándares informáticos estructurados, basados en las buenas prácticas de las tecnologías de Información y en patrones de calidad internacionales, que proporcionen las pautas, guías, y soportes necesarios para la identificación y documentación de un sistema de información. Se

diversifican en: **i) marco de trabajo informático**, que define un conjunto estandarizado de fases, actividades, tareas, herramientas e instrumentos que serán utilizados en la resolución de situaciones genéricas de riesgo informático, de optimización sistémica o de mejora financiera, mediante la creación de un modelo de indicadores y competencias inherentes al área de estudio. Igualmente contempla la generación de enfoques de desarrollo de sistemas, así como la concepción de buenas prácticas para el uso y aplicación de tecnologías de la Información (TI) y **ii) estandarización de procesos sistémicos**, que se refiere a la documentación integral de procesos operativos, relacionados directamente con el uso de las TI dentro de la organización. Persigue la elaboración y regulación de manuales de procedimientos sistémicos y políticas de operatividad, amparados bajo normas y certificaciones internacionales, a fin de mantener modelos de soporte trabajos adaptables, constantes y documentados.

III) La auditoría, que se orienta hacia el control y evaluación de los departamentos, documentos y procesos sistémicos de una organización, estableciendo criterios objetivos e independientes, a fin de determinar el cumplimiento eficaz de las funciones y su concantenación con las normas establecidas. La finalidad de la esta línea potencial, es la búsqueda de posibles inconsistencias en el objeto de estudio, y la proposición de soluciones basadas en estándares informáticos eficaces y confiables. Se diversifica en: **i) métodos para evaluación**, que se refiere a la instauración de actividades, tácticas e instrumentos novedosos, que aseguren la eficacia en la auditoría de procesos de gestión de servicios, comunicaciones, relaciones con el cliente, manejo de la información y toma de decisiones para la gerencia en la organización. Incluye el diseño de técnicas de recolección de datos, formatos de verificación, mecanismos de control y estrategias de función operativa y **ii) auditoría de seguridad informática**, que comprende los métodos de autenticación y verificación de los sistemas de información, así como el estudio de la consistencia y seguridad de los modelos informáticos de la organización. Se contempla tanto la elaboración de programas (*software*) auditores como la

gestión de pruebas de cumplimiento para la selección de anomalías dentro de cualquier sistema automatizado.

1.7. Cibernética

Esta línea de investigación contempla el conjunto de estudios tecnológicos, electrónicos y mecánicos que actúan en conjunción para conformar sistemas que imitan la conducta de los seres vivientes, mediante el control dirigido de dispositivos, maquinarias o programas. Busca la transformación continua de información, por lo cual se relaciona con los conceptos de inteligencia artificial, sistemas expertos y automatización, con la finalidad de servir de soporte al hombre en acciones que superan sus capacidades naturales o buscan el aumento de la comodidad.

La cibernética también se aplica al estudio de la psicología, la inteligencia artificial, los servomecanismos, la economía, la neurofisiología, la ingeniería de sistemas y al de los sistemas sociales. La palabra cibernética ha dejado de identificar un área independiente de estudio y la mayor parte de la actividad investigadora se centra ahora en el estudio y diseño de redes neuronales artificiales. Su objetivo es diseñar máquinas capaces de reaccionar y operar con más precisión y rapidez que los seres vivos, dado que su interés fundamental es diseñar máquinas o hetero-organizaciones, donde alguien organiza el sistema. La cibernética contempla de igual forma los sistemas de comunicación y control de los organismos vivos que los de las máquinas. Para obtener la respuesta deseada en un organismo humano o en un dispositivo mecánico, habrá que proporcionarle, como guía para acciones futuras, la información relativa a los resultados reales de la acción prevista. Los intereses principales para esta línea de investigación son:

1) **La domótica**, que se refiere a la automatización de los procesos del hogar, mediante la utilización de diversos dispositivos tecnológicos que sirven como coadyuvantes a la seguridad, confort, y reutilización de la energía para su posterior aprovechamiento.

Se diversifica en: *i) sistemas inalámbricos*, que se refiere al conjunto de dispositivos que emiten y reciben señales, ondas y datos sin necesidad de conexión física, así como también a los componentes de software que se encargan de facilitar ésta transmisión de información, con el fin de mejorar tanto las condiciones de transmisión como el aprovechamiento de espacio, convirtiéndolo en el sistema de transferencia más ergonómico de la actualidad; *ii) monitoreo*, que se describe al uso de un sistema que constantemente monitoriza una acción, para recibir y procesar las señales de información en busca de un problema o una amenaza, para así reflejar una señal mediante alertas. Monitoriza una red de computadoras en busca de componentes defectuosos o lentos, para luego informar a los administradores de redes mediante correo electrónico, mensajes, teléfonos, u otras alarmas. Es un subconjunto de funciones de la administración de redes y *iii) biométrica*, que es el estudio de métodos automáticos para el reconocimiento único de humanos basados en uno o más rasgos conductuales o físicos intrínsecos. La biometría informática es la aplicación de técnicas matemáticas y estadísticas sobre los rasgos físicos o de conducta de un individuo, para verificar identidades o para identificar individuos.

Los Intereses principales para esta línea de investigación son: *domótica (sistemas inalámbricos, monitoreo, biométrica), automatización de procesos industriales (controladores lógicos, sistemas de adquisición de datos y controles industriales).*

II) La automatización de procesos industriales, que se define como sistemas abiertos que controlan dispositivos electromecánicos en una organización, por medio del trabajo en conjunto de componentes electrónicos y un software controlador abocado a estandarizar los movimientos y procesos de todo el sistema. La automatización de procesos va más allá de la mecanización de un proceso físico, más bien se enfoca en la sustitución de seres humanos en procesos industriales, para la optimización de los resultados en las áreas de supervisión, auditoría de procesos y cálculos estadísticos de producción. Se diversifica en: *i) controladores lógicos programables,*

comúnmente conocidos como PLC, que son dispositivos con un microprocesador de aplicación específica para el control de procesos industriales. Contempla tanto la unidad de control como la interfaz que regula las operaciones generales del mecanismo, teniendo en cuenta el enfoque interdisciplinario que debe darse para la unión electrónico-lógica; **ii) sistemas de control, supervisión y adquisición de datos**, denominados SCADA, son sistemas concebidos para la inspección y revisión de procesos a distancia, contemplando el puente de comunicación entre los controladores lógicos programables y programas dedicados, con la intervención del enfoque sistémico hombre-máquina, proporcionando información del proceso de campo a distintos usuarios del sistema cibernético, a través de controles autónomos digitales o autómatas programables; y los **iii) controles industriales**, que son un conjunto de dispositivos electromecánicos y electrónicos controlados por un programa para medir y supervisar magnitudes físicas como velocidad o frecuencia de aparatos mecánicos.

1.8. Sistemas de comunicaciones

Esta línea de investigación está orientada hacia la planificación, diseño, instalación, configuración, mantenimiento y resolver problemas de las redes de datos actuales. De igual manera, garantizar la planificación de la seguridad de redes y transmisión de información y el funcionamiento. Su objetivo es dar apoyo a todas las áreas y procesos de la organización, con acciones para facilitar el cumplimiento del plan estratégico, alinear los procesos y propiciar un ambiente positivo, dado que todas las operaciones efectuadas a lo largo de los procesos, lo que permite no sólo conocer on-line todo lo que sucede, sino también disponer de una herramienta fundamental para la toma de decisiones y la corrección de posibles desviaciones producidas. Los intereses principales para esta línea de investigación son:

I) Las redes de datos, que contempla las infraestructuras o redes de comunicación que se ha diseñado específicamente a la transmisión de información mediante el intercambio de datos. Las redes de

datos se diseñan y construyen en arquitecturas que pretenden servir a sus objetivos de uso, generalmente, están basadas en la conmutación de paquetes y se clasifican de acuerdo a su tamaño, la distancia que cubre y su arquitectura física. Esto abarca: **la seguridad en redes**, que se encarga de mantener bajo protección los recursos y la información con que se cuenta en la red, a través de procedimientos basados en una política de seguridad tales que permitan el control de lo actuado; **la virtualización**, que es la tecnología utilizada para crear una versión virtual de un dispositivo o recurso, como un servidor, un dispositivo de almacenamiento, una red o incluso un sistema operativo, donde se divide el recurso en uno o más entornos de ejecución. Existen virtualizaciones tales como virtualización de almacenamiento, virtualización de servidor, virtualización a nivel sistema operativo, virtualización de aplicación y virtualización de red; **la alámbrica**, que contempla diseño, instalación y configuración de dispositivos de red en una organización para trabajos como: impresión, compartir archivos, programas administrativos, cámara de seguridad IP, respaldo de Información e internet; y **la inalámbrica**, que abarca el diseño, instalación, configuración de dispositivos de red en una organización para trabajos como: impresión, compartir archivos, programas administrativos, cámara de seguridad IP, respaldo de información e internet.

II) **La transmisión de datos**, que está centrada en la transmisión de información codificada, de un punto a uno o más puntos, mediante señales eléctricas, ópticas, electroópticas o electromagnéticas; con la finalidad de reducir tiempo y esfuerzo, aumentar la velocidad de entrega de la información, reducir costos de operación, aumentar la calidad y cantidad de la información.

Actividades del capítulo I

1. ¿Cuál debe ser el papel del ingeniero de sistemas en la sociedad actual?
2. ¿Cuál debe ser la orientación del ingeniero de sistemas en la sociedad actual?
3. ¿En qué áreas debe estar capacitado el ingeniero de sistemas para dar respuestas a los nuevos retos del mundo globalizado?
4. ¿De qué forma debe asumir ingeniero de sistemas su rol de investigador, desarrollador e innovador de nuevas tecnologías?
5. ¿En dónde comenzó la investigación en ingeniería de sistemas y en donde ha desembocado actualmente?
6. ¿Cuáles son las líneas de investigación que se desarrollan actualmente en ingeniería de sistemas?
7. ¿En qué consisten los sistemas de procesamiento de información y cuál es la diversificación de sus líneas investigativas?
8. ¿En qué consisten los sistemas inteligentes y cuál es la diversificación de sus líneas investigativas?
9. ¿En qué consisten los sistemas y arquitectura de procesos y cuál es la diversificación de sus líneas investigativas?
10. ¿En qué consisten los métodos y estandarización de sistemas y cuál es la diversificación de sus líneas investigativas?
11. ¿En qué consisten la cibernética y cuál es la diversificación de sus líneas investigativas?
12. ¿En qué consisten los sistemas de comunicaciones y cuál es la diversificación de sus líneas investigativas?

Metodología investigativa en ingeniería de sistemas

2.1. Fundamentos metodológicos de una investigación

La metodología investigativa aplicada por el ingeniero de sistemas debe comenzar por el diseño de una estructura teórica que va a representar la “*columna vertebral de un proyecto de investigación*” debido a que ésta, le da el soporte teórico a los resultados y conclusiones obtenidas para que los mismos tengan una “*validez objetiva y científica*”. Es decir, se obtenga un “*conocimiento nuevo*” a través del planteamiento formal de una “*investigación*”, la cual entenderemos como una actividad o diligencia humana, cuyo objetivo u orientación es la “*obtención o descubrimiento de nuevos conocimientos*” y su aplicación para la solución a problemas o interrogantes de carácter científico. Una “*investigación científica*” es un metódico y sofisticado proceso de obtener “*conocimiento científico*” cuyos resultados provienen de la aplicación del “*método científico*” para resolver o tratar explicar determinados fenómenos. La investigación en concreto comienza con la determinación del problema a investigar y casi siempre surge en forma de pregunta en la mente del investigador. Un problema de investigación es una situación confusa que amerita una explicación cuya respuesta sólo se podrá obtener después de realizada la investigación.

El “*método científico*” se puede definir como una armazón conformada por conjunto sistemático de reglas, principios y criterios de acción, que dirigen el proceso de investigación con el fin de dar respuesta a una o varias interrogantes, exponer y/o confirmar una teoría

o crear un nuevo conocimiento; cuyo objetivo es la objetividad en que la ciencia avance al verdadero conocimiento de las cosas. Lo anterior va a consistir en el recorrido por una serie de etapas a recorrer para obtener un “conocimiento válido” desde el punto de vista científico minimizando la subjetividad. Estas etapas son: *la observación, las preguntas, las hipótesis, la experimentación, la teoría y ley*. “*La observación*” que consiste en fijar la atención en una porción del universo para identificar realidades o acontecimientos específicos del cosmos a través de los sentidos. “*Las preguntas*” que surgen una vez realizada la observación, producto de la curiosidad del investigador respecto a una realidad o fenómeno observado. Las preguntas deben comenzar con adjetivos tales como: un “*qué*”, un “*cómo*”, un “*dónde*”, un “*cuándo*”; un “*cuál es*”, un “*cuales son*”, o un “*para qué es*”. *Por ejemplo: ¿Cuál es la causa por la cual las plantas son de color verde?*

“*Las hipótesis*” que representan afirmaciones que dan una solución preliminar a un problema y surgen cuando se dan una o más respuestas lógicas a las preguntas planteadas por el investigador. Cada hipótesis es una declaración o respuesta tentativa a una pregunta de investigación y ésta puede ser verdadera o falsa, por tanto debe ser sometida a comprobación teórica o experimental. Los resultados de comprobación teórica o experimental, determinarán el carácter final “*verdadero o falso*” de la hipótesis. *Por ejemplo: Las plantas son de color verde porque tienen un pigmento que refleja ese color.* “*La experimentación*” se realiza cuando las hipótesis o soluciones preliminares a un problema son sometidas a pruebas sistemáticas para comprobar su ocurrencia.

La experimentación consiste determinar la “*causa y/o efecto*” de un fenómeno a partir de controlar las condiciones de un proceso a través de “*dos muestras aleatorias similares*”, llámese a una “*muestra aleatoria experimental*” (es aquella a la que se aplica el estímulo) y llámese a otra “*muestra aleatoria de control*” (es aquella a la que no se le aplica el estímulo) a partir de la definición de un conjunto de “*variables independientes (estímulos o causas) y variables dependientes (efectos)*”. Las muestras aleatorias “*experimental y control*”, son so-

metidas a las mismas condiciones, excluyendo el “estímulo o variable independiente” que se aplica a la muestra aleatoria experimental.

“Las conclusiones” a partir de la experimentación la hipótesis original es evaluada y se determina si es verdadera o falsa, evaluándose los resultados en base de su “alcance espacio/temporal” y de acuerdo a eso, se puede concluir si se ha llegado una teoría o a una ley. “Una teoría” es una declaración parcialmente verdadera o totalmente verdadera, que se verifica por medio de la experimentación y/o de las evidencias, y que sólo es válida para un espacio y un determinado. Como ilustración se puede decir que *la clorofila hace que la pigmentación de las plantas sea verde y le ayuda a producir energía mediante la fotosíntesis.*

“Una ley” es una declaración parcialmente verdadera o totalmente verdadera, que se verifica como verdadera por medio de la experimentación y/o de las evidencias en todo tiempo y lugar. Es permanente e inmutable y es comprobable en cualquier tiempo y espacio. Si una teoría se verificara como verdadera en todo tiempo y lugar, entonces es considerada como ley. Como ilustración se puede decir que *la ley de gravitación universal es una ley física clásica presentada por Isaac Newton que describe la interacción gravitatoria entre distintos cuerpos con masa, pues ocurre en todo tiempo y lugar del universo conocido.*

2.2. Diseño y contextualización de una investigación

El diseño y contextualización de una investigación se refiere a la bosquejo y delineación esquemática estratégica de la investigación desde la que se soporta el investigador para dar respuesta a las interrogantes, dificultades o inconveniente planteados en el problema objeto de un estudio, ya que el diseño y contextualización de una investigación va a indicar a través de un mecanismo secuencial las formas de plantear los pasos a seguir para alcanzar los objetivos de la investigación y los métodos de cómo lograr dar respuesta a las interrogantes o hipótesis del problema planteado en la investigación;

es por tanto necesario que previo a la selección del diseño de investigación, se tengan claros los objetivos de la investigación. Los aspectos fundamentales a plasmarse en el diseño de una investigación, están divididas en tres aristas o dimensiones, estas son: *la naturaleza de la investigación*, que representa la esencia y las propiedades intrínsecas que caracterizan la investigación; *el nivel de profundidad de la investigación*, que va a determinar la graduación o calado que quiere realizar en una investigación; y *el paradigma metodológico de la investigación*, que se refiere a la modalidad del estudio que se va a realizar, el cual orienta su metodología de acuerdo a la finalidad general del estudio. Es decir, se refiere a la estrategia que adopta el investigador para responder a las interrogantes del problema y las dificultades o inconveniente del mismo. En la Figura 2, se expone un cuadro que esquematiza el diseño de una investigación:

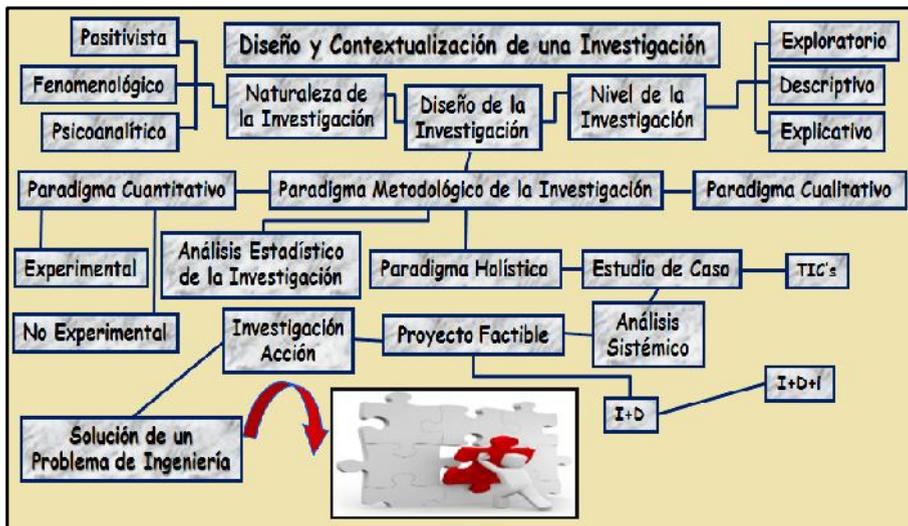


Figura 2. Diseño y Contextualización de una Investigación. Fuente: Autores.

La primera arista o dimensión en el **“diseño de una investigación”** es **“la naturaleza de la investigación”**, que como se mencionó, representa la esencia y las propiedades intrínsecas que caracterizan la investigación. En un sentido bastante general, existen tres concepciones donde el investigador pudiera ubicar la naturaleza de su in-

investigación: “*investigación de naturaleza positivista*”, “*investigación de naturaleza fenomenológica*” e “*investigación de naturaleza psicoanalítica*”. La investigación de *naturaleza positivista* busca conocer las causas que producen los fenómenos, generalmente estableciendo una relación experimental “*causa-efecto*”, con independencia de los estados subjetivos a través del “*método científico*”. En la concepción positivista se considera que la formulación de un problema es adecuada, siempre y cuando exprese una relación entre dos o más variables que generalmente se denominan variable dependiente y variable independiente.

Para la concepción positivista el proceso metódico de producción de conocimiento científico es: empírico, analítico, hipotético y deductivo. Una pregunta generadora de la investigación en la concepción positivista sería: ¿cuál es el efecto de los videos didácticos asíncronos colgados en una plataforma virtual de enseñanza-aprendizaje que complementan los materiales digitales: textos/guías de los contenidos programáticos de un curso virtual de ingeniería? El objetivo podría ser demostrar que el uso de videos didácticos asíncronos en la enseñanza de un curso virtual de ingeniería mejora el aprendizaje de los contenidos programáticos.

La investigación de *naturaleza fenomenológica* busca conocer la “*esencia de los fenómenos*”, y el conocimiento estricto de los mismos tal y como se muestran u ofrecen a la conciencia, apelando a la experiencia intuitiva. En la fenomenología o “*ciencia de fenómenos*”, predomina el empleo de métodos cualitativos de investigación, en la descripción de los fenómenos depuradas de elementos empíricos, que interpretan la realidad mediante la reducción. La esencia significa los aspectos invariantes tales como *formas objetivas* (*es la cualidad de lo objetivo, perteneciente o relativo al objeto en sí mismo, con independencia de la propia manera de pensar o de sentir*), *estructuras subjetivas* (*es la construcción personal del objeto*), a partir de actitudes o vivencias. Una pregunta generadora de la investigación en la concepción fenomenológica sería: ¿cuál es el desempeño laboral de los padres de familias constituidas en los barrios pobres de Bogotá?

o ¿cuáles son los problemas de adaptación y los cambios en la vida familiar de los inmigrantes de Colombia radicados en EEUU? Los objetivos podrían ser: i) Desarrollar una literatura comparativa sobre estudios minuciosos del desempeño laboral, de acuerdo al nivel de formación escolar, de los padres de familias constituidas en los barrios pobres de Bogotá o ii) Comprobar y refinar el concepto de inmigración desde una perspectiva cultural popular de pobreza. Y la investigación de *naturaleza psicoanalítica* busca comprender los fenómenos desde una postura referida a los “*estratos más profundos y ocultos del psiquismo o mente humana*”; lo que otorga a los datos nuevos y más complejos niveles de significación. El investigador y sus relaciones con los sujetos involucrados en la investigación a través la obtención de significados de experiencias, comportamientos y vivencias, que van a servir para dar respuestas a las preguntas formuladas en torno a un determinado objeto de estudio. Su estrategia metodológica es la producción cualitativa de datos a través de “*mecanismos conversacionales flexibles y abiertos*”, donde los sujetos involucrados en la investigación se vean implicados de manera activa a lo largo de todo el proceso. Una pregunta generadora de la investigación en la concepción psicoanalítica sería: ¿cómo las instituciones y los acontecimientos políticos se reflejan en la vida inconsciente del estudiante en cuanto a sus proyecciones de desempeño laboral? El objetivo podría ser: Incorporar una perspectiva psicosocial al estudio de los estudiantes provenientes de estratos populares.

La segunda arista o dimensión en el “*diseño de una investigación*” es “*el nivel de profundidad de la investigación*”, que como de menciono, va a determinar la graduación o calado que quiere realizar en una investigación. Esta se subdivide en: “*investigación de carácter exploratorio*”, “*investigación de carácter descriptivo*” e “*investigación de carácter explicativo*”. La investigación de *carácter exploratorio*, es entendida como aquella que se efectúa sobre un fenómeno u objeto desconocido o poco estudiado, por lo que sus resultados constituyen una “*visión aproximada de un objeto o fenómeno*”, es decir, un nivel superficial de conocimientos del objeto o fenómeno.

Ejemplo 1: Plantee un estudio exploratorio en el “uso de recursos digitales en el aprendizaje de las matemáticas” en los estudiantes cursantes del básico de ingeniería de la *Universidad Simón Bolívar* en donde se exponga a través de una matriz de datos la información recabada y se realice un análisis exploratorio.

Supóngase que se emplearon tres herramientas de recolección de información tales como *entrevista, fuentes secundarias de información y encuesta*, se pudo acceder a la información contenida en una muestra física conformada por 25 estudiantes cursantes del básico de ingeniería de la *Universidad Simón Bolívar* respecto al uso de “recursos digitales en el aprendizaje de las matemáticas o variables” tales como: *videos, archivos de texto, presentaciones en power point, paginas interactivas y otros medios digitales*; de donde se extrajo la información para construir la “matriz de datos” mostrada en la Tabla 4.

Se organizan los “datos o resultados” expuestos en la “matriz de datos” obtenida a partir de realizar entrevistas, utilizar fuentes secundarias de información y realizar encuestas; a través de una “tabla de porcentajes” para dar una descripción inicial del fenómeno de aprendizaje de matemáticas en red (Tabla 5).

Para ilustrar gráficamente la información contenida en la Tabla, se puede utilizar un “Gráfico de doble barra” que permita diferenciar las opciones “lo utiliza” o “no lo utiliza” para cada uno de los “recursos digitales” en el aprendizaje de la matemática (Figuras 3 y 4).

El estudio exploratorio realizado, concluye que el “uso de recursos digitales en el aprendizaje de las matemáticas” en los estudiantes cursantes del básico de ingeniería de la *Universidad Simón Bolívar* está dividido en 5 tipos y los más utilizados por los estudiantes cursantes del básico de ingeniería de la *Universidad Simón Bolívar*, son los “archivos de texto” con un 96% de aceptación y los “archivos de video” con un 68% de aceptación. Le siguen los archivos de “presentaciones en power point” con 40% de aceptación. Y los menos usados fueron “otros medios digitales de enseñanza” con un 32% de aceptación y las “páginas interactivas de enseñanza” con un 28%, tal y como lo ilustra la Figura 5.

Tabla 4. Uso de recursos digitales en el aprendizaje de las matemáticas (1)

Muestra de Estudiantes	Y(Columnas): Variables Analizadas				
	Videos	Archivos de Texto	Presentaciones Power Point	Páginas Interactivas	Otros Medios Digitales
Estudiante 1	Si	Si	Si	Si	Si
Estudiante 2	Si	Si	No	No	Si
Estudiante 3	Si	Si	Si	No	No
Estudiante 4	No	Si	No	No	No
Estudiante 5	Si	Si	No	No	No
Estudiante 6	Si	Si	Si	No	No
Estudiante 7	No	Si	No	No	No
Estudiante 8	No	No	No	No	Si
Estudiante 9	Si	Si	No	No	No
Estudiante 10	No	Si	Si	No	No
Estudiante 11	Si	Si	No	No	No
Estudiante 12	Si	Si	No	Si	No
Estudiante 13	Si	Si	No	No	No
Estudiante 14	Si	Si	Si	Si	Si
Estudiante 15	Si	Si	No	No	Si
Estudiante 16	Si	Si	Si	No	No
Estudiante 17	Si	Si	No	No	No

X (Fila): Elementos Tomados de la Población

Tabla 4 (Continuación)

	Muestra de Estudiantes	Y(Columnas): Variables Analizadas				
		Videos	Archivos de Texto	Presentaciones Power Point	Páginas Interactivas	Otros Medios Digitales
X (Fila): Elementenos Tomados de la Población	Estudiante 18	Si	Si	Si	No	No
	Estudiante 19	Si	Si	Si	Si	Si
	Estudiante 20	No	Si	No	No	No
	Estudiante 21	No	Si	No	No	Si
	Estudiante 22	No	Si	Si	No	No
	Estudiante 23	Si	Si	No	Si	No
	Estudiante 24	Si	Si	Si	Si	No
	Estudiante 25	No	Si	No	Si	Si

Fuente: Autores.

Tabla 5. Uso de recursos digitales en el aprendizaje de las matemáticas (2)

Recurso Digital	Lo utiliza	No lo utiliza	Porcentaje que lo utiliza	Porcentaje que no lo utiliza
Videos	17	8	68%	32%
Archivos de Texto	24	1	96%	4%
Presentación Power Point	10	15	40%	60%
Paginas Interactivas	7	18	28%	72%
Otros Medios Digitales	8	17	32%	68%

Fuente: Autores.

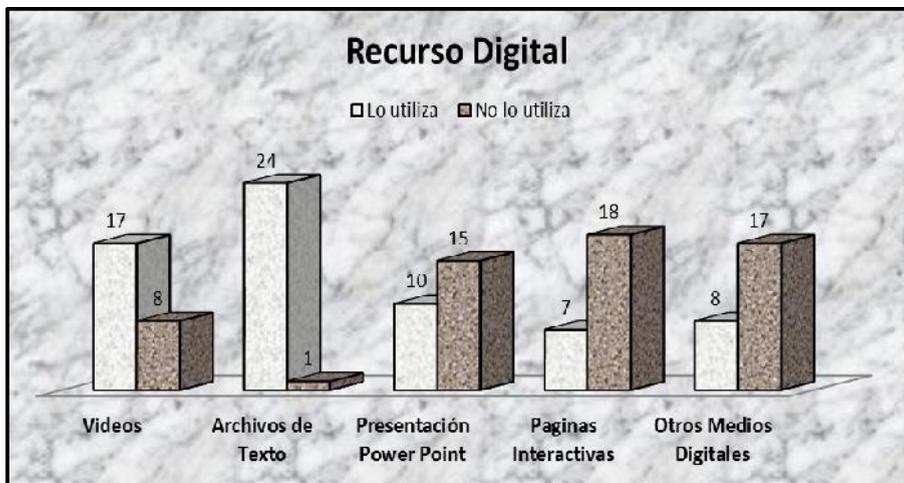


Figura 3. Visualización de los resultados de recursos digitales. Fuente: Autores.

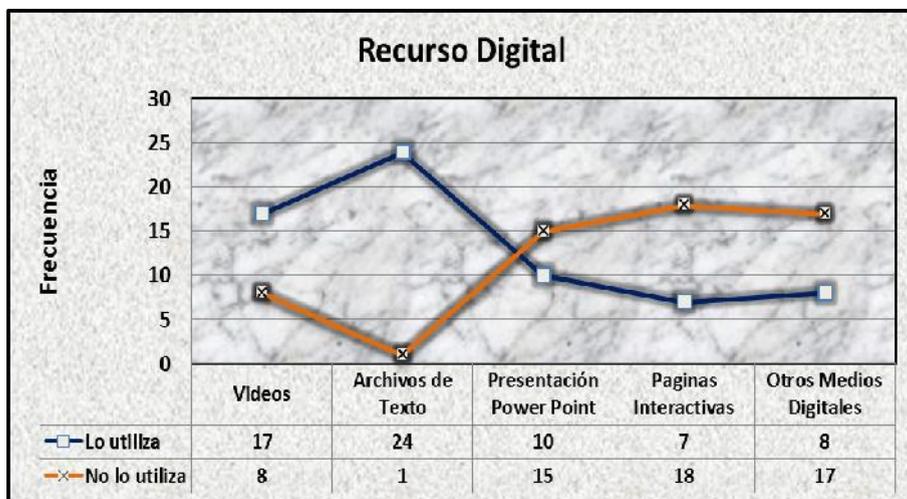


Figura 4. Uso de recursos digitales en el aprendizaje de las matemáticas
Fuente: Autores.

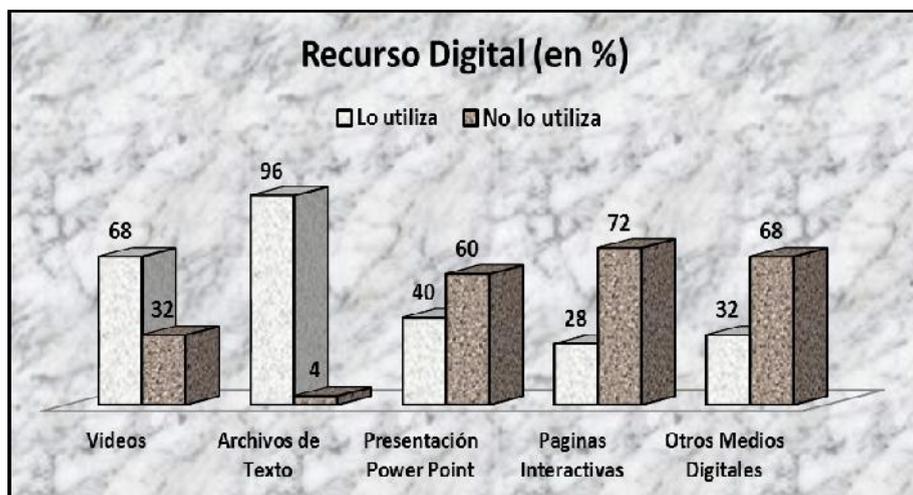


Figura 5. Porcentaje de uso de recursos digitales en el aprendizaje de las matemáticas. Fuente: Autores.

La investigación de *carácter descriptivo*, es entendida como aquella que consiste en la descripción de las características de un fenómeno individual o grupal, de ahí que radique en “*describir la estructura o comportamiento de un fenómeno*”. Los resultados de este tipo de investigación se ubican en un nivel intermedio en cuanto a la profundidad de los conocimientos se refiere.

Ejemplo 2: Plantee un estudio descriptivo en el “*uso de recursos digitales en el aprendizaje de las matemáticas*” en los estudiantes cursantes del básico de ingeniería de la *Universidad Simón Bolívar* en donde se exponga a través de una matriz de datos la información recabada y además se realice un análisis descriptivo.

Supóngase que a través de la realización inicial de un estudio exploratorio se concluyó que el “*uso de recursos digitales en el aprendizaje de las matemáticas*” en los estudiantes cursantes del básico de ingeniería de la *Universidad Simón Bolívar* está dividido en 5 tipos y los más utilizados por los estudiantes cursantes del básico de ingeniería de la *Universidad Nacional Experimental del Táchira*, son los “*archivos de texto*” y los “*archivos de video*”; le siguen los “*archivos en power point*” (*percepción y estructura y lugar donde ocurre el fenómeno*). Dado lo anterior, se pretende realizar un estudio descriptivo en el “*lapso de tiempo de una semana*” (*tiempo*), en el que se describa el uso (*propiedades*) y compare (*interrelacione*): “*la frecuencia de uso*”, “*el tiempo de conexión en el uso*” y “*el lugar de conexión*”, de los *archivos de texto, video y de presentaciones en power point*. Supóngase que a través del empleo de dos herramientas de recolección de información tales como “*la entrevista*” y “*el test*”, se pudo acceder a la información contenida en una muestra física conformada por 27 estudiantes cursantes del básico de ingeniería de la *Universidad Simón Bolívar* respecto al uso de “*recursos digitales en el aprendizaje de las matemáticas o variables*” tales como: *archivos de texto, videos y presentaciones en power point*; de donde se extrajo la información para construir la “*matriz de datos*” mostrada en la Tabla 6.

Tabla 6. Frecuencia de uso de recursos digitales en el aprendizaje de las matemáticas (1)

		Y (variable analizada): uso de recursos en archivos de texto (T), uso de recursos archivos de videos (V) y uso de recursos en archivos de presentación en Power Point (P)																	
Muestra de estudiantes	Frecuencia semanal:			Tiempo de conexión:			Lugar de conexión:												
	A (De 0 a 1 vez), B (De 2 a 3 veces), C (4 o más veces)			A (De 0 a 1 hora), B (De 2 a 3 horas), C (4 o más horas)			A (En casa), B (En la Universidad), C (En otro lugar)												
X: Fila Elementenos Tomados de la Población	Estudiante 1	T: A	V: C	P: A	T: B	V: C	P: A	T	A	B	Ø	V	A	B	C	P	A	Ø	Ø
	Estudiante 2	T: B	V: B	P: B	T: C	V: B	P: B	T	A	B	C	V	A	B	C	P	A	B	C
	Estudiante 3	T: C	V: C	P: B	T: C	V: C	P: C	T	A	B	C	V	A	Ø	C	P	Ø	B	Ø
	Estudiante 4	T: B	V: B	P: B	T: B	V: C	P: A	T	A	B	Ø	V	A	B	Ø	P	A	B	Ø
	Estudiante 5	T: C	V: C	P: A	T: B	V: C	P: A	T	Ø	B	Ø	V	Ø	B	Ø	P	Ø	B	Ø
	Estudiante 6	T: C	V: C	P: C	T: C	V: C	P: B	T	A	Ø	C	V	A	Ø	C	P	A	Ø	C
	Estudiante 7	T: C	V: B	P: B	T: C	V: C	P: B	T	Ø	Ø	C	V	Ø	Ø	C	P	Ø	Ø	C
	Estudiante 8	T: A	V: C	P: A	T: B	V: C	P: A	T	A	B	Ø	V	A	B	C	P	A	Ø	Ø
	Estudiante 9	T: B	V: B	P: B	T: C	V: A	P: B	T	A	B	C	V	A	B	C	P	A	B	C
	Estudiante 10	T: C	V: C	P: B	T: C	V: C	P: C	T	A	B	C	V	A	Ø	C	P	Ø	B	Ø
	Estudiante 11	T: B	V: B	P: B	T: B	V: C	P: A	T	A	B	Ø	V	A	B	Ø	P	A	B	Ø
	Estudiante 12	T: C	V: C	P: A	T: B	V: C	P: A	T	Ø	B	Ø	V	Ø	B	Ø	P	Ø	B	Ø
	Estudiante 13	T: C	V: C	P: C	T: C	V: C	P: B	T	A	Ø	C	V	A	Ø	C	P	A	Ø	C

Tabla 6 (Continuación)

		Y (variable analizada): uso de recursos en archivos de texto (T), uso de recursos archivos de videos (V) y uso de recursos en archivos de presentación en Power Point (P)																			
Muestra de estudiantes	Frecuencia semanal: A (De 0 a 1 vez), B (De 2 a 3 veces), C (4 o más veces)	Tiempo de conexión:						Lugar de conexión:													
		A (De 0 a 1 hora),		B (De 2 a 3 horas),		C (4 o más horas)		A (En casa),		B (En la Universidad),		C (En otro lugar)									
X: Fila Elementenos Tomados de la Población	Estudiante 14	T: C	V: B	P: B	T: C	V: C	P: B	T	Ø	Ø	C	V	Ø	Ø	C	P	Ø	Ø	C		
	Estudiante 15	T: B	V: B	P: B	T: B	V: C	P: A	T	A	B	Ø	V	A	B	Ø	P	A	B	Ø		
	Estudiante 16	T: C	V: C	P: A	T: A	V: C	P: A	T	Ø	B	Ø	V	Ø	B	Ø	P	Ø	B	Ø		
	Estudiante 17	T: C	V: C	P: C	T: C	V: C	P: B	T	A	Ø	C	V	A	Ø	C	P	A	Ø	C		
	Estudiante 18	T: C	V: B	P: B	T: C	V: C	P: B	T	Ø	Ø	C	V	Ø	Ø	C	P	Ø	Ø	C		
	Estudiante 19	T: A	V: C	P: A	T: A	V: C	P: A	T	A	B	Ø	V	A	B	C	P	A	Ø	Ø		
	Estudiante 20	T: B	V: B	P: B	T: C	V: B	P: B	T	A	B	C	V	A	B	C	P	A	B	C		
	Estudiante 21	T: C	V: C	P: B	T: C	V: C	P: C	T	A	B	C	V	A	B	C	P	Ø	B	Ø		
	Estudiante 22	T: B	V: A	P: B	T: A	V: C	P: A	T	A	B	Ø	V	A	B	Ø	P	A	B	Ø		
	Estudiante 23	T: C	V: C	P: A	T: B	V: C	P: A	T	Ø	B	Ø	V	Ø	B	Ø	P	Ø	B	Ø		
	Estudiante 24	T: B	V: B	P: B	T: B	V: C	P: A	T	A	Ø	C	V	A	Ø	C	P	A	Ø	C		
	Estudiante 25	T: C	V: C	P: A	T: A	V: C	P: A	T	Ø	Ø	C	V	Ø	Ø	C	P	Ø	Ø	C		
	Estudiante 26	T: C	V: C	P: C	T: C	V: C	P: B	T	A	B	Ø	V	A	B	Ø	P	A	B	Ø		
Estudiante 27	T: C	V: A	P: B	T: C	V: C	P: B	T	Ø	B	Ø	V	Ø	B	Ø	P	Ø	B	Ø			

Fuente: Autores.

Se organizan los datos a través de una tabla de frecuencias (Tabla 7). En esta tabla las variables son los: archivos de texto, archivos de videos y archivos de presentaciones en power point. La información se extrajo de una muestra aleatoria de “27 estudiantes de ingeniería” cursantes del básico de ingeniería de la Universidad Simón Bolívar en el lapso de tiempo de una semana.

Tabla 7. Frecuencia de uso de recursos digitales en el aprendizaje de las matemáticas (2)

Recursos Digitales	Frecuencia Semanal: A: (De 0 a 1 vez), B: (De 2 a 3 veces), C: (4 o más veces)			Tiempo de Conexión Semanal A: (De 0 a 1 hora), B: (De 2 a 3 horas), C: (4 o más horas)			Lugar de Conexión Semanal A: (En Casa), B: (En La Universidad), C: (En Otro Sitio)		
	A	B	C	A	B	C	A	B	C
Archivos de Texto	3	8	16	4	9	14	18	19	14
Archivos de Video	2	10	15	1	2	24	18	17	17
Archivos en Power Point	8	15	4	13	11	3	15	16	11

Fuente: Autores.

Luego, para ilustrar gráficamente la información contenida en la tabla, se puede utilizar un “*gráfico de doble barra*” que permita diferenciar las opciones “*lo utiliza*” o “*no lo utiliza*” para cada uno de los “*recursos digitales*” en el aprendizaje de la matemática.

La investigación de *carácter explicativo*, es entendida como aquella que se encarga de buscar el porqué de los hechos mediante el establecimiento de la “*relación causa-efecto de un fenómeno*”. En este sentido, los estudios explicativos pueden ocuparse tanto de la determinación de las causas (*investigación post facto*), como de los efectos (*investigación experimental*), mediante la “*prueba de hipóte-*

sis” a través de la creación de un “*experimento del fenómeno*”. Sus resultados y conclusiones constituyen el nivel más profundo de conocimientos. De ahí que, en función del alcance de una investigación, esta puede ser “*exploratoria, descriptiva o explicativa*”. Estos tipos de investigación se consideran como etapas cronológicas de todo estudio científico y cada una tiene una finalidad diferente: primero se “*explora*” un tema para conocerlo mejor, luego se “*describen*” las variables involucradas, después “*explican*” un fenómeno al buscar relaciones “*causa-efecto*” intentado explicar la influencia de unas variables sobre otras en términos de efecto o causalidad.

Ejemplo 3: Plantee un estudio explicativo para exponer el funcionamiento de un objeto (*supóngase desconocido*) llamado “Dado”.

Se comienza con el planteamiento de un experimento para poder realizar un estudio del fenómeno. Su objetivo es la obtención de resultados a través de la experimentación con el objeto en estudio. Luego se procede a analizar una muestra de resultados obtenidos con la experimentación con el objeto. A los resultados que se generan a partir de la experimentación con el objeto, se les llama “*datos u observaciones*”. El listado total de resultados o datos que arroja el experimento es el tamaño “*n*” de la “*muestra aleatoria*”. El objetivo de la investigación explicativa es la organización, descripción y análisis de los datos obtenidos.

Supóngase que al experimentar con el objeto en estudio llamado “Dado”, se obtuvieron los resultados expuestos en la Tabla 8.

A través de la estadística organizan las “*observaciones*” (*resultados o datos*) para entender (*cuantificar*) el funcionamiento del fenómeno en estudio. En este caso un “Dado” (Tabla 9).

Tabla 8. Frecuencia de resultados del lanzamiento de un dado (1)

Muestra física de tamaño 23 (n=23): Lanzamientos del dado	X: Marca de la cara superior (Característica) $(X(a) = x)$
Lanzamiento 1: a_1	$X(a_1) = x_1 = 5$
Lanzamiento 2: a_2	$X(a_2) = x_2 = 4$
Lanzamiento 3: a_3	$X(a_3) = x_3 = 3$
Lanzamiento 4: a_4	$X(a_4) = x_4 = 2$
Lanzamiento 5: a_5	$X(a_5) = x_5 = 1$
Lanzamiento 6: a_6	$X(a_6) = x_6 = 1$
Lanzamiento 7: a_7	$X(a_7) = x_7 = 5$
Lanzamiento 8: a_8	$X(a_8) = x_8 = 4$
Lanzamiento 9: a_9	$X(a_9) = x_9 = 4$
Lanzamiento 9: a_{10}	$X(a_{10}) = x_{10} = 3$
Lanzamiento 9: a_{11}	$X(a_{11}) = x_{11} = 2$
Lanzamiento 9: a_{12}	$X(a_{12}) = x_{12} = 1$
Lanzamiento 9: a_{13}	$X(a_{13}) = x_{13} = 6$
Lanzamiento 9: a_{14}	$X(a_{14}) = x_{14} = 6$
Lanzamiento 9: a_{15}	$X(a_{15}) = x_{15} = 3$
Lanzamiento 9: a_{16}	$X(a_{16}) = x_{16} = 3$
Lanzamiento 9: a_{17}	$X(a_{17}) = x_{17} = 5$
Lanzamiento 9: a_{18}	$X(a_{18}) = x_{18} = 5$
Lanzamiento 9: a_{19}	$X(a_{19}) = x_{19} = 6$
Lanzamiento 9: a_{20}	$X(a_{20}) = x_{20} = 2$
Lanzamiento 9: a_{21}	$X(a_{21}) = x_{21} = 4$
Lanzamiento 9: a_{22}	$X(a_{22}) = x_{22} = 6$
Lanzamiento 9: a_{23}	$X(a_{23}) = x_{23} = 4$

Nota: La expresión " $X(a) = x$ " se lee "Resultado de la Cara Superior"

Fuente: Autores.

Tabla 9. Frecuencia de resultados del lanzamiento de un dado (2)

Clases	Resultado de la cara superior ($X(a) = x$)	Frecuencia (f_i)
Clase 1	1	$f_1 = 3$ (el 1 se repite 3 veces)
Clase 2	2	$f_2 = 3$ (el 2 se repite 3 veces)
Clase 3	3	$f_3 = 4$ (el 3 se repite 4 veces)
Clase 4	4	$f_4 = 5$ (el 4 se repite 5 veces)
Clase 5	5	$f_5 = 4$ (el 5 se repite 4 veces)
Clase 6	6	$f_6 = 4$ (el 6 se repite 4 veces)

Fuente: Autores.

$$\sum_{i=1}^6 f_i = 23 = n \longrightarrow \text{Número de veces que se realizó el experimento}$$

De lo anterior, se puede concluir que el objeto arroja seis resultados, estos resultados son: $\{1, 2, 3, 4, 5, 6\}$ y que la frecuencia con que ocurren estos resultados no es muy diferente. Para obtener conclusiones más confiables, se ha de realizar el experimento “muchas veces” tabulando y analizando los resultados obtenidos. Esto permitirá la construcción un “modelo matemático” (una ecuación) que cuantifique y explique los resultados arrojados por el fenómeno en estudio.

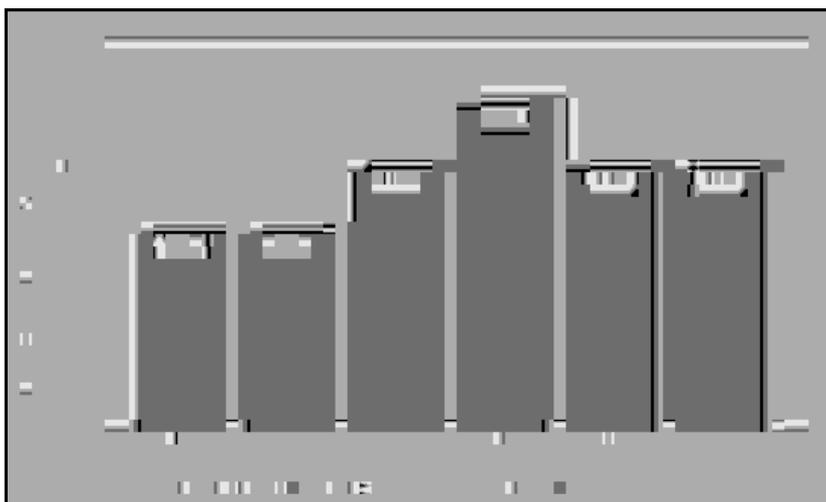


Figura 6. Frecuencia de resultados del lanzamiento de un dado. Fuente: Autores.

La tercera arista o dimensión en el “*diseño de una investigación*” es “*el paradigma metodológico de la investigación*”, que como de menciono, se refiere a la modalidad del estudio que se va a realizar, el cual orienta su metodología de acuerdo a la finalidad general del estudio. Es decir, se refiere a la estrategia que adopta el investigador para responder a las interrogantes del problema y las dificultades o inconveniente del mismo. Esta se subdivide en: *el paradigma cuantitativo, el paradigma cuantitativo y el paradigma holístico.*

El paradigma cuantitativo, lo determina una metodología de investigación “*desde afuera (observa)*” que construye un conocimiento de forma “*deductiva*”; es decir: *objetiva, estructurada, secuencial, lógica y sistemática*, que busca establecer cuantitativamente leyes generales a partir del método científico aplicado a generalizaciones empíricas en ciencias naturales, en donde: *se experimenta, se controla y se mide*, apoyado en técnicas de análisis estadístico inferencial de datos. El método cuantitativo persigue un análisis y explicación lo más exacta posible de lo que ocurre. El paradigma cuantitativo está orientado a “*productos o resultados*”, a generalizar una realidad estable y a dar confiabilidad de “*datos o resultados duros*”. El paradigma cualitativo, lo determina una metodología de investigación “*desde adentro (interactúa)*” que construye un conocimiento de forma “*inductiva*”: *subjetiva, no estructurada, simbólica, dialéctica y flexible*, que busca comprender significados que se infieren de las acciones individuales y conductas sociales en las ciencias humanas, en donde: *se comprende, se interpreta y se explica*, apoyado en técnicas de análisis estadístico descriptivo de datos que cuentan con significados a través de la interpretación humana. El método cualitativo investiga los ¿por qué?, los ¿cómo?, los ¿dónde? y los ¿cuándo? De ahí, que el método cualitativo se utilicen muestras pequeñas enfocadas a un tema en particular. La metodología cualitativa se basa en principios teóricos como la fenomenología, la hermenéutica, la interacción social utilizando métodos de recolección tales como matrices de información que no pueden ser plasmadas en números. La idea es explorar las relaciones sociales y describir la realidad tal como la experi-

mentan los protagonistas. El *paradigma cualitativo*, está orientado: a una realidad dinámica, no generalizable, orientada al descubrimiento exploratorio, descriptivo e inductivo. Los paradigmas cuantitativo y cualitativo no son excluyentes y comúnmente se integran en un planteamiento holístico del investigador a fin de poder establecer triangulaciones metodológicas. Una de las principales diferencias entre la investigación cualitativa y la investigación cuantitativa es que la primera “*busca explicar las razones del fenómeno*”, la cuantitativa “*busca probar mediante datos numéricos*” porque esas razones pueden llegar a ser ciertas. Y el *paradigma holístico*, podría decirse que el adjetivo “*holística*” más bien hace referencia a una actitud del investigador hacia el proceso de generación del conocimiento. Una actitud de apertura y de búsqueda de una comprensión integradora de los paradigmas cualitativos y cuantitativos en el estudio de un fenómeno. En este sentido, investigación holística alude a una comprensión integradora de la investigación y propone generar un modelo del proceso investigativo que recogiera los aportes de los diferentes modelos investigativos, a manera de estadios u holón (o *estado jerárquico*) que incluye, las capacidades, patrones y funcionamiento de la etapa anterior (*de los holones previos*) como una especie de círculos concéntricos o estadios (*figuras dentro de figuras*) llámese espiral holística. Los “*estadios u holones*” están conformados en una propuesta, por diez modelos investigativos que van desde los más simples, hasta los de orden más elevados, mayor complejidad y organización a medida que se sube por cada “*vuelta de la espiral, estadio u holón de conocimiento*”, estos modelos son: *investigación exploratoria, investigación descriptiva, investigación analítica, investigación comparativa, investigación explicativa, investigación predictiva, investigación proyectiva, investigación interactiva, investigación confirmatoria e investigación evaluativa*. La espiral holística; pretende ser aplicable tanto a las ciencias sociales como a las ciencias naturales y proporciona al método científico unos diseños o estadios jerárquicos; donde cada tipo de investigación tiene un sistema coherente de análisis que responde a la realidad en estudio. La estructura

holística aparece cuando las partes se unifican, donde ningún holón parece ser más importante o dominante, y cada uno contribuye de forma más o menos igualitaria al rescate de la totalidad. En cada uno de esos estadios se desarrollan diferentes procesos metodológicos y van proporcionando resultados necesarios (*bajo la forma de conocimiento nuevo*), para el logro del objetivo general. Desde una comprensión holística el ser humano se aproxima al conocimiento en un proceso permanente en “*espiral*” donde cada resultado alcanza grados de complejidad cada vez más avanzados.

2.3. Metodologías investigativas cuantitativas

La investigación experimental, es un proceso que consiste en someter a un objeto o grupo de individuos, a determinadas condiciones, estímulos o tratamiento (*variable independiente*), para observar los efectos o reacciones que se producen (*variable dependiente*). En este tipo de investigación se delimita y somete el objeto y las variables de estudio a condiciones controladas y conocidas por el investigador, para observar los resultados que cada variable ejerce sobre el objeto bajo estudio. Sus elementos básicos: experimento, variable independiente, variable dependiente, variables de control, hipótesis, muestra de control y muestra experimental. Se caracteriza por que la variable experimental no está comprobada y su objeto de estudio depende del investigador. Sus etapas se pueden sintetizar en: delimitar y definir el objeto de la investigación o problema, plantear una hipótesis de trabajo, elaborar el diseño del experimento, realizar el experimento, analizar los resultados, obtener conclusiones.

El diseño experimental es aquel según el cual el investigador manipula una variable experimental no comprobada, bajo condiciones estrictamente controladas. Su objetivo es describir de qué modo y porque causa se produce o puede producirse un fenómeno. La investigación experimental es netamente explicativa, por cuanto su propósito es demostrar que los cambios en la variable dependiente fueron causados por la variable independiente, estableciéndose con precisión una “*relación causa-efecto*”. Los experimentos puros, en

la investigación experimental, se considera como puro a aquel que cumple con tres condiciones: *i)* Se acepta la manipulación intencional de una o más variables independientes. La variable independiente es considerada como supuesta causa en una relación entre variables; es la condición antecedente, y al efecto provocado por dicha causa se le denomina variable dependiente (*consecuente*).

Un experimento se lleva a cabo para analizar si una o más variables independientes afectan a una o más variables dependientes y por qué lo hacen. En un auténtico experimento, la variable independiente resulta de interés para el investigador por ser la variable que se hipotetiza, que será una de las causas que producen el efecto supuesto. Para obtener respuesta de esta relación causal supuesta, el investigador manipula la variable independiente y observa si la dependiente varía o no, *“manipular es hacer variar o dar distintos valores a la variable independiente”*. La variable dependiente se mide, la variable dependiente no se manipula, sino que se mide para ver el efecto de que la manipulación de la variable independiente tienes de ella. A la presencia de la variable independiente se le llama *“tratamiento experimental”* o *“estímulo experimental”* y se aplica a la *“muestra experimental”* y la no presencia o ausencia de la variable independiente se aplica a la *“muestra de control”*. Luego los dos grupos son comparados para ver si el grupo que fue expuesto a la variable independiente difiere del grupo que no fue expuesto, *ii)* Se puede medir el efecto que la variable independiente tiene en la variable dependiente. Esto es igualmente importante y como en la variable dependiente se observa el efecto, la medición debe ser válida y confiable. Si no podemos asegurar que se midió adecuadamente, los resultados no servirán. En la planeación de un experimento se debe precisar cómo se van a manipular las variables independientes y cómo a medir las dependientes, y *iii)* Se logra tener el control o validez interna de la situación experimental.

El término *“control”* consiste en que si en el experimento se observa que una o más variables independientes hacen variar a las dependientes, la variación de estas últimas se deba a la manipulación

y no a otros factores o causas; si se observa que una o más independientes no tienen efecto sobre las dependientes, se pueda estar seguro de ello. En términos coloquiales, “control” significa saber qué está ocurriendo realmente con la relación entre las variables independientes y las dependientes. Cuando hay control podemos conocer la relación causal. Lograr “control” en un experimento es controlar la influencia de otras variables extrañas en las variables dependientes, para que así podamos saber realmente si las variables independientes tienen o no efecto en la dependiente.

Los cuasi experimentos, son aquellos en donde en donde “el investigador no puede hacer la asignación al azar de los sujetos a los grupos experimental y de control”. La falta de aleatorización introduce posibles problemas de validez interna y externa. Sin embargo, se puede controlar: cuándo llevar a cabo las observaciones, cuándo aplicar la variable independiente o tratamiento y cuál de los grupos recibirá el tratamiento. En los diseños cuasi-experimentales también manipulan deliberadamente al menos una variable independiente para ver su efecto y relación con una o más variables dependientes, solamente que difieren de los experimentos verdaderos en el grado de seguridad o confiabilidad que pueda tenerse sobre la equivalencia inicial de los grupos. La interpretación es similar, las comparaciones son las mismas y los análisis estadísticos iguales. El término cuasi significa casi por lo que un diseño cuasi-experimental casi alcanza el nivel de experimental, el criterio que le falta para llegar a este nivel es que no existe ningún tipo de aleatorización, es decir, no hay manera de asegurar la equivalencia inicial de los grupos experimental y control. Se toman grupos que ya están integrados por lo que las unidades de análisis no se asignan al azar ni por pareamiento aleatorio. La carencia de aleatorización implica la presencia de posibles problemas de validez tanto interna como externa. La validez interna se ve afectada por el fenómeno de selección, la regresión estadística y el proceso de maduración. La validez externa se ve afectada por la variable población, es decir, resulta difícil determinar a qué población pertenecen los grupos. En los diseños cuasi-experimentales el expe-

rimentador no puede hacer la asignación al azar de los sujetos a los grupos experimentales y de control. Y si puede controlar: cuándo llevar a cabo las observaciones, cuándo aplicar la variable independiente o tratamiento y cuál de los grupos recibirá el tratamiento. Aunque estos diseños no garantizan un nivel de validez interna y externa como en los experimentales, ofrece un grado de validez suficiente, lo que hace muy viable su uso en el campo de la educación y de la psicología. Los principales pasos en el desarrollo de un experimento o cuasi experimento son: **i)** decidir cuántas variables independientes y dependientes deberán ser incluidas en el experimento o cuasi experimento, **ii)** elegir los niveles de manipulación de las variables independientes y traducirlos en tratamientos experimentales, **iii)** desarrollar el instrumento o instrumentos para medir la(s) variable(s) dependiente(s), **iv)** seleccionar una muestra de personas para el experimento (idealmente representativa de la población), **v)** reclutar a los sujetos del experimento o cuasi experimento. Esto implica tener contacto con ellos, darles las explicaciones necesarias e indicarles el lugar, día, hora y persona con quien deben presentarse. Siempre es conveniente darles el máximo de facilidades para que acudan al experimento, **vi)** seleccionar el diseño experimental o cuasi experimental apropiado para muestras, hipótesis, objetivos y preguntas de investigación, **vii)** planear cómo vamos a manejar a los sujetos que participen en el experimento. Es decir, elaborar una ruta crítica de qué van a hacer los sujetos desde que llegan al lugar del experimento hasta que se retiran (paso a paso), **viii)** en el caso de experimentos “verdaderos”, dividirlos al azar o emparejarlos; y en el caso de cuasi experimentos analizar cuidadosamente las propiedades de los grupos intactos, **ix)** aplicar las pre pruebas (cuando las haya), los tratamientos respectivos (cuando no se trate de grupos de control) y las post pruebas.

La investigación no experimental, se usa cuando no es factible utilizar un diseño experimental verdadero que cumpla con los requerimientos rigurosos del modelo experimental básico, es decir, no se construye una situación específica si no que se observa las que exis-

ten. El investigador no ejerce control ni manipulación alguna sobre las variables de estudio y se realiza sin manipular en forma deliberada ninguna variable. Se observan los hechos tal y como se presentan en su contexto real y en un tiempo determinado o no, para luego analizarlos. Es un método de control parcial, *“basado en la identificación de los factores que pueden intervenir en la validez interna y externa de un fenómeno”*. Se incluye el uso de grupos intactos de sujetos para la realización del experimento, puesto que en un estudio no siempre es posible seleccionar objetos al azar, esta no delimita y ni somete el objeto y las variables de estudio a condiciones controladas y conocidas por el investigador. Por ejemplo: *“Si se estudia el efecto del consumo de alcohol cuando la madre está embarazada, sabemos que el alcohol hace daño a los embriones y en un diseño estrictamente experimental implicaría que a las madres se les fuese aleatoriamente asignado beber alcohol. Esto sería ilegal por el posible daño que el estudio podría causar a los embriones, para observar los resultados que cada variable ejerce sobre el objeto bajo estudio, de ahí que debería tomar muestras estratificadas en las que se sepa de antemano el hecho de que la madre consumió o no alcohol en el periodo de embarazo”*. Las etapas de este tipo de investigación son: **i) Definir el problema, ii) Revisar la literatura, iii) Enunciar hipótesis, iv) Describir los supuestos en que se basan las hipótesis, v) Determinar los procedimientos para:** Seleccionar los sujetos o unidades de observación, Determinar cuáles técnicas va a utilizar en la recolección de los datos, Probar si estas técnicas son confiables, esto es, si producen la información deseable, **vi) Determinar procedimientos para analizar los datos, tales como pruebas estadísticas de asociación y de significación, vii) Recoger datos y viii) Describir, analizar e interpretar los resultados en términos claros y precisos.**

La investigación de campo, la investigación de campo llamada también *“investigación sobre el terreno”*, porque se aboca a estudiar los fenómenos en la realidad misma donde se producen y corresponde a un tipo de diseño de investigación que se basa en informaciones obtenidas directamente de la realidad, sin modificar su ambiente natural. Los estudios de campo permiten indagar *“in situ”* (en el sitio o

en el lugar), los efectos de la interrelación entre los diferentes tipos de variables sociológicas, psicológicas, educacionales, etc. La investigación de campo se realiza a través de un proceso sistemático, riguroso y racional de recolección, tratamiento, análisis y presentación de datos, basado en una estrategia de recolección directa de la realidad de las informaciones necesarias para la investigación. La investigación de campo consiste en la recolección de datos directamente de la realidad donde ocurren los hechos, sin manipular o controlar las variables. Estudia los fenómenos sociales en su ambiente natural. El investigador no manipula variables debido a que esto hace perder el ambiente de naturalidad en el cual se manifiesta.

La investigación descriptiva, se caracteriza por que no hay manipulación de variables, estas se observan y se describen tal como se presentan en su ambiente natural. La investigación descriptiva pretende llegar a conocer las situaciones, costumbres y actitudes predominantes a través de la descripción exacta de las actividades, objetos, procesos y colectivos. Su meta no se limita a la recolección de datos, sino a la predicción e identificación de las relaciones que existen entre dos o más variables. Los investigadores no son meros tabuladores, sino que recogen los datos sobre la base de una hipótesis o teoría, exponen y resumen la información de manera cuidadosa y luego analizan minuciosamente los resultados, a fin de extraer generalizaciones significativas que contribuyan al conocimiento. Sus etapas son: **i) se examinan las características del problema escogido, ii) se define el problema y se formula la hipótesis, enunciando los supuestos en que se basan las hipótesis y los procesos adoptados, iii) se eligen los temas y las fuentes apropiados, iv) se seleccionan o elaboran técnicas para la recolección de datos, v) se establecen, a fin de clasificar los datos, categorías precisas, que se adecuen al propósito del estudio y permitan poner de manifiesto las semejanzas, diferencias y relaciones significativas, vi) se verifican la validez de las técnicas empleadas para la recolección de datos, vii) se realizan observaciones objetivas y exactas, viii) se describen, analizan e interpretan los datos obtenidos, en términos claros y precisos.**

La investigación descriptiva correlacional, se caracteriza comúnmente por que estudia las relaciones entre las variables dependientes y las variables independientes, es decir, se mide el grado y nivel de relación existente entre las dos variables. Este tipo de investigación se utiliza para determinar la medida en que dos variables se correlacionan entre sí, es decir si “*el grado en que las variaciones que sufre un factor se corresponden con las que experimenta el otro*”. Las variables pueden hallarse estrecha o parcialmente relacionadas entre sí, pero también es posible que no exista entre ellas relación alguna. Puede decirse, en general, que la magnitud de una correlación depende de la medida en que los valores de dos variables aumenten o disminuyan en la misma o en diferente dirección. Si los valores de dos variables aumentan o disminuyen de la misma manera, existe una correlación positiva; si, en cambio, los valores de una variable aumentan en tanto que disminuyen los de la otra, se trata de una correlación negativa; y si los valores de una variable aumentan, los de la otra pueden aumentar o disminuir, entonces hay poca o ninguna correlación. En consecuencia la gama de correlaciones se extiende desde la perfecta correlación negativa hasta la no correlación o la perfecta correlación positiva. Las técnicas de correlación son muy útiles en los estudios de carácter predictivo. Si bien el coeficiente de correlación sólo permite expresar en términos cuantitativos el grado de relación que dos variables guardan entre sí, no significa que tal relación sea de orden causal. Para interpretar el significado de una relación se debe recurrir al análisis lógico, porque la computación estadística no dilucida el problema. Sus riesgos son los mismos que en los estudios causales comparativos.

La investigación descriptiva causal comparativa, se caracteriza por que además de pretender describir cómo es un fenómeno, también se quiere saber de qué manera y por qué ocurre, entonces se comparan semejanzas y diferencias que existen entre fenómenos, para descubrir los factores o condiciones que parecen acompañar o contribuir a la aparición de ciertos hechos y situaciones. Por la complejidad y naturaleza de los fenómenos sociales, es menester

estudiar las relaciones de causalidad. Este tipo de estudio se usa en los casos en que los investigadores no pueden manejar una variable independiente y establecer los controles requeridos en los experimentos. En un estudio causal comparativo el investigador analiza la situación vital en la cual los sujetos han experimentado el fenómeno que se quiere investigar. Después de estudiar las semejanzas y diferencias que hay entre dos situaciones, entonces podrá describir los factores que parecen explicar la presencia del fenómeno en una situación y su ausencia en la otra. Esta investigación establece que *“si dos o más instancias del fenómeno investigado tienen sólo una circunstancia en común, en la cual todas las instancias concuerdan, es la causa (o efecto) del fenómeno dado”*. Este método proporciona al investigador la doble posibilidad de control sobre sus conclusiones acerca de las relaciones de causalidad. Las dificultades posibles de explicar los fenómenos en este tipo de estudios reside en la imposibilidad de establecer un control más allá de poner a prueba tantas hipótesis alternativas como sea posible; si al estudiar el problema el factor produce un efecto determinado no se incluye entre los puntos considerados, entonces no será posible averiguar la causa (si se desea hallar las posibles causas de los fenómenos y desechar los factores aleatorios se debe poseer suficiente información general acerca de tales fenómenos y elaborar cuidadosamente sus procedimientos de observación); al exigir que sea un solo factor el que determine la aparición o ausencia de un fenómeno, muchas veces en los fenómenos sociales complejos se obedece a múltiples causas; cuando dos variables se hallan relacionadas entre sí es difícil determinar cuál de ellas es la causa y cuál el efecto; al intentar clasificar a los sujetos en grupos dicotómicos a fin de establecer comparaciones entre ellos, los fenómenos sociales sólo presentan similitudes si los incluimos en las más amplias categorías, aunque sabemos que los hechos sociales no se clasifican automáticamente en categorías exclusivas (por lo general la comparación entre dos variables indefinidas proporciona escasa información útil sobre el fenómeno que se pretende explicar); cuando se trata de estudios en los que se comparan situaciones

normales, la tarea de seleccionar no requiere los mismos cuidados y precauciones que en el caso de los estudios experimentales, generalmente resulta difícil hallar grupos de elementos que sean similares en todos sus aspectos, excepto en lo que respecta al hecho de hallarse expuestos a una variable distinta (siempre existe el peligro de que los grupos presenten diferencias en relación con otras dimensiones –salud, inteligencia, antecedentes familiares, experiencia anterior– que pueden afectar los resultados del estudio).

La investigación pre experimental, se caracteriza por que se aplica en una sola variable que no es manipulada, en donde se tiene un mínimo grado de control del experimento, se realiza con fines únicamente exploratorios y se realiza sin comparación con grupos utilizándose muestras poco representativas. El proceso pre experimental consiste en: **i)** administrar el estímulo o variable independiente, **ii)** medir el efecto o variable dependiente, y **iii)** realizar un análisis exploratorio de los resultados obtenidos. Requiere una validez interna mínima (*pruebas de que el diseño de un experimento o estudio tuvo algún efecto sobre las observaciones y los resultados; es decir, si el mismo midió la característica deseada*) para aceptar la aplicación del modelo y respecto a la validez externa (proceso de generalización de los resultados obtenidos de un grupo de muestra pequeño, a menudo en un entorno de laboratorio, a toda la población) existe poca posibilidades de generalizar resultados por la ausencia de control experimental de las variables y de la hipótesis. *La investigación ex post facto*, se caracteriza por que se realiza “sobre hechos cumplidos” es apropiada para establecer posibles relaciones de “causa-efecto” observando que ciertos hechos han ocurrido y buscando en el pasado los factores que los hayan podido ocasionar. Es análoga en su metodología a la investigación no-experimental. Se diferencia del verdadero experimento en que en éste “la causa se introduce en un momento determinado y el efecto se viene a observar algún tiempo después”. Por ejemplo: **i)** identificar las características de los estudiantes que obtuvieron altas notas en los estudios del semestre que transcurrió, **ii)** determinación de factores asociados al cáncer pul-

monar tales como el hábito de fumar tabaco o el tipo de ocupación que hayan tenido las personas que padecen esa enfermedad, **iii)** determinación de la relación entre las características de una campaña política, la situación del país y el éxito en unas elecciones presidenciales. En la investigación ex post facto no se controlan las variables independientes, dado que el estudio se basa en analizar eventos ya ocurridos de manera natural. Como el evento ya ha ocurrido los métodos de análisis pueden ser descriptivos o experimentales. Las investigaciones no experimentales de acuerdo al período o lapso de tiempo que se tome para realizar la investigación se clasifican en: diseños transeccionales o transversales y diseños longitudinales.

La investigación transeccional o transversal, se caracteriza por que son estudios diseñados para medir o determinar la prevalencia (# de casos favorables / # de elementos de la población en estudio) de una o varias características de interés de una población en un punto específico del tiempo. Se centra en analizar cuál es el nivel o estado de una o varias variables en un momento dado. Es como tomar una fotografía de algo que sucede. Por ejemplo: **i)** Investigar el número de empleados, desempleados y subempleados en una ciudad en cierto momento; **ii)** Determinar el nivel de escolaridad de los trabajadores de una industria en un punto en el tiempo; **iii)** Analizar la relación entre las ventas y la capacidad de compra (en determinado momento); **iv)** Analizar si hay diferencias en el contenido nutricional de un mismo producto de tres fábricas diferentes. Los diseños de investigación transeccionales pueden dividirse en dos: **descriptivos y correlacionales causales**. Resumiendo: **i)** se realiza en un lapso de tiempo corto, **ii)** es como tomar una instantánea de un evento.

La investigación transversal descriptiva, se caracteriza porque son estudios diseñados para medir la existencia o prevalencia de una o varias características de interés de una población en un punto específico del tiempo y proporcionar su descripción, aquí se establecen hipótesis descriptivas. Por ejemplo: **i)** estudio diseñado para determinar la prevalencia de HIV en mujeres embarazadas residentes en el área de captación de un centro de salud rural en el mes de Enero de 2016, **ii)**

la realización de una encuesta nacional de opinión sobre las tendencias de los votantes durante periodos antes de las elecciones, **iii**) la realización de una encuesta sobre la percepción de características de un producto en el mercado por sus clientes para determinar sus preferencias. En ciertas ocasiones el investigador pretende hacer descripciones comparativas entre grupos o subgrupos de personas, objetos o indicadores (esto es, en más de un grupo). Por ejemplo: un investigador que deseara describir el nivel de aceptación de una bebida láctea en tres ciudades.

La investigación transversal correlacional causal, se caracteriza por que tiene como objetivo determinar relaciones causales entre dos o más variables en un momento determinado. Se trata también de descripciones, pero no de variables individuales sino de sus relaciones, sean éstas puramente correlacionales o relaciones causales. En estos diseños lo que se mide es la relación causal entre variables en un tiempo determinado. Una investigación que pretendiera indagar la relación entre la calidad de un producto cárnico y las ventas a durante una época del año, observando qué tan relacionadas están ambas variables (*se limita a ser correlacional*). Una investigación que estudiase el cómo la motivación intrínseca (*propia*) influye en la productividad de los trabajadores de línea de grandes empresas industriales, de determinado país y en cierto momento, observando si los obreros más productivos son los más motivados, y en caso de que así sea, evaluando el por qué y cómo es que la motivación intrínseca contribuye a incrementar la productividad (*esta investigación establece primero la correlación y luego la relación causal entre las variables correlacionadas*). Los diseños transeccionales correlacionales causales buscan describir correlaciones entre variables o relaciones causales entre variables, en uno o más grupos de personas u objetos o indicadores en un momento determinado. Como ilustración se puede plantear la recolección simultánea de los resultados de potenciales en aquellas personas expuestas a cada factor de riesgo con aquellas no expuestas al factor de riesgo, de población definida. Luego se compara la prevalencia del resultado en aquellas personas

expuestas a cada factor de riesgo con la prevalencia del resultado en aquellos no expuestos. Por ejemplo: *En el estudio de HIV en mujeres embarazadas se podría recolectar simultáneamente información sobre distintos factores de riesgo: número de compañeros sexuales, uso de preservativos, uso de drogas intravenosas, historia de otras enfermedades de transmisión sexual, transfusiones de sangre, etc.*

La investigación longitudinal, se caracteriza porque son estudios diseñados para analizar cambios a través del tiempo de una o varias características de interés de una población. Estos recolectan datos en diversos puntos o periodos especificados de tiempo, para hacer inferencias respecto al cambio: “*evolución y/o desarrollo y de tendencia*”. Por ejemplo: **i)** *se diseña un estudio para analizar cómo evolucionan los niveles de consumo de leche durante cinco años en una ciudad, ii)* *se diseña un estudio para analizar cómo ha cambiado el contenido de proteína de la leche durante un determinado periodo: un año dos años o los últimos tres años en una región.* Los diseños longitudinales recolectan datos sobre las variables (*características*) de una población o las relaciones entre estas características en dos o más instantes de tiempo para evaluar los cambios. Resumiendo: *La investigación longitudinal hace estudios y análisis en un tiempo prolongado dividido en varios cohortes tomados en diferentes periodos de tiempo viendo la evolución o la tendencia de la característica bajo estudio, es como ver una película de un evento.*

La investigación longitudinal de tendencia, se caracteriza por que son estudios diseñados para analizar los cambios a través del tiempo en determinadas variables o en las relaciones entre éstas observando la tendencia de la característica bajo estudio. Por ejemplo: *el análisis de la tendencia ideológica o política de una población o sub-población en diferentes lapsos de tiempo (o cohortes tomados en diferentes periodos de tiempo) previos a las elecciones.*

La investigación longitudinal evolutivo y/o desarrollo, se caracteriza porque son estudios diseñados para examinar los cambios a través del tiempo en determinadas variables o en las relaciones entre éstas observando la evolución y/o desarrollo de la característica

bajo estudio en subpoblaciones o grupos específicos. Por ejemplo: la evolución y/o desarrollo de características microbiológicas de sabor y textura de un producto vinícola a través del tiempo (o cohortes tomados en diferentes periodos de tiempo).

2.4. Metodologías investigativas cualitativas

La investigación hermenéutica, es entendida como la metodología de investigación documental que realiza una interpretación o comprensión de los hechos apoyándose en el relato. A partir de la hermenéutica se realizaron y se realizan, por ejemplo, el análisis de textos bíblicos. En este sentido, la hermenéutica procura comprender los textos a partir del ejercicio interpretativo intencional, contextual y en gran medida se trata de traspasar las fronteras contenidas en la “*física de la palabra*” para lograr la captación del sentido de éstas en tanto plasmadas en un papel. La hermenéutica no se preocupa tanto por la intención del actor, sino que toma la acción como una vía para interpretar el contexto social de significado más amplio en el que está inmerso. La hermenéutica parte de que el ser humano es interpretativo por naturaleza y que el círculo de comprensión e interpretación es infinito. Las observaciones que se registran en el diario de campo para su análisis posterior.

La investigación etnográfica es entendida como la metodología que se enfoca en la descripción y estudio del modo de vida de: un grupo de individuos, de una unidad social concreta, pudiendo ser ésta una familia o un claustro de profesores de una escuela, durante un cierto período, utilizando la observación participante o las entrevistas para conocer su comportamiento social mediante un estudio detallado que incluye su cultura, lenguaje y costumbres desde la perspectiva del investigador/participante. Se describe e interpreta los fenómenos sociales “*desde dentro*”. Los mecanismos de recolección de la información son a través de observación registrada un diario de campo (cuaderno con fecha, hora, lugar, observaciones,...) para su análisis posterior y las entrevistas.

La investigación etnometodológica, es entendida como la metodología de investigación que está interesada en las formas en que las personas, interactúan unas con otras en situaciones cotidianas a través del análisis conversacional (*diálogos cotidianos*) con el propósito de ver la realidad social. La etnometodología enfatiza la necesidad de apartar las propias creencias e ideas sobre la estructura social y de enfocar primordialmente en la forma como los miembros de un grupo, manejan, reproducen y logran un sentido de la estructura social. Se describe e interpreta los fenómenos sociales “*desde dentro*”. Los mecanismos de recolección de la información son a través de observación registrada un diario de campo para su análisis posterior.

La investigación documental, es entendida como una “*estrategia metodológica de obtención de información*”, que supone por parte del investigador el instruirse acerca de la realidad objeto de estudio a través de documentos de diferente materialidad (escritos: cartas, libros, audio-visuales, numéricos,...), con el fin de acreditar las justificaciones e interpretaciones que realiza en el análisis y reconstrucción de un fenómeno que tiene características de historicidad. La investigación documental, a través de la observación y el análisis de la “*documentación*” nos permite volver la mirada hacia un tiempo pasado para de este modo comprender e interpretar una realidad actual (*sincrónica*) a la luz de acontecimientos pasados que han sido los antecedentes que han derivado en los consecuentes de situaciones, acontecimientos y procesos de una realidad determinada. La investigación documental es un proceso basado en la búsqueda, recuperación, análisis, crítica e interpretación de datos secundarios, es decir, los obtenidos y registrados por otros investigadores en fuentes documentales: impresas, audiovisuales o electrónicas. Este tipo de investigación, se fundamenta en la revisión sistemática, rigurosa y profunda del material documental de cualquier clase y procura el análisis de los fenómenos o el establecimiento de la relación entre dos o más variables.

La Investigación biográfica, es una metodología de investigación que muestra el testimonio subjetivo de una persona en el que

se recogen los acontecimientos y las valoraciones que esta persona hace de su propia existencia. La investigación biográfica busca habilidades en el investigador de: observar, escuchar, comparar y escribir. Una de las vertientes del método biográfico es el relato de vida que consiste en una historia de una vida tal y como la persona que la ha vivido la cuenta. La investigación biográfica permite explorar la dinámica de situaciones concretas a través de relatos, la cual se materializa en una historia de vida e historia oral. Se basa en narraciones autobiográficas generadas en el diálogo interactivo de una entrevista. Esta metodología se utiliza para preservar el conocimiento de eventos históricos. Se utilizan biogramas (*registros biográficos de carácter más sucinto y que suponen la recopilación de una amplia muestra de biografías personales a efectos comparativos*). En este método se emplean documentos personales que poseen valor afectivo o simbólico para el sujeto, autobiografías, diarios personales, correspondencia, fotos, videos y objetos personales obtenidos por el investigador a través de las entrevistas. El método biográfico consta de: **i) Etapa inicial** (*se elabora el planteamiento teórico, la hipótesis de partida, se justifica la elección del método, se delimita el universo y se explicitan los criterios de selección del o los informantes por biografiar*), **ii) Transcripción y registro** (*se obtiene toda la información biográfica, se registra y se transcribe, esta puede ser original o literal, cronológico, personal o temático*), **iii) Análisis e interpretación**, **iv) Presentación y publicación de los relatos biográficos** (*estos pueden ser historias de vida de relato único, relatos de vida paralelos o relatos de vida cruzados o polifónicos*). Las técnicas empleadas son: la entrevista y el análisis documental de manera profunda, los documentos personales y el registro biográfico.

2.5. Metodologías investigativas alternativas

El estudio de caso, es un tipo de investigación que se centra en el estudio de situaciones o entidades sociales únicas, llamadas casos, que merecen interés de investigación. Esta metodología es muy útil para estudiar problemas prácticos o situaciones determinadas y es

uno de los más utilizados por el trabajo social. Como ilustración se pueden mencionar: *Las deficiencias en el suministro eléctrico. Caso: Las comunas del centro, El nivel de uso de herramientas telemáticas. Caso: Los de primeros semestres de la Universidad de Pamplona, La convivencia colocada en su contexto social, económico y político. Caso: Comuna Antonia Santos, Las influencias de una página web orientada a denunciar los problemas con el hampa. Caso: Comuna Aniversario Uno, La enseñanza virtual de la matemática versus la enseñanza presencial de la matemática. Caso: Estudiantes de ingeniería de la Universidad de Simón Bolívar.* La función del estudio de caso es fomentar el análisis para comprender fenómenos conceptos teóricos, que ayuden a comprender o solucionar el problema. Se estudia un caso desde una forma particularista, descriptiva, heurística (método para aumentar el conocimiento) e inductiva. Al final del estudio de caso se encontrará el registro del caso, donde se expone éste de forma descriptiva, con cuadros, imágenes, recursos narrativos, etc. Es necesario precisar que el estudio de casos puede incluir tanto estudios de un solo caso como de múltiples casos (según sea una o varias las unidades de análisis) pero su propósito fundamental es comprender la particularidad del caso, en el intento de conocer cómo funcionan todas las partes que los componen y las relaciones entre ellas para formar un todo. Las etapas a considerar en un estudio de caso son: **i) evaluación de la situación para determinar si se dan las condiciones para llevar a cabo la investigación, ii) delimitación del problema, definiéndolo y concentrándose en los aspectos que más relevantes, iii) recopilación de la información y acomodándolos en un orden de importancia, iv) análisis de hechos que permita tener las conclusiones para la toma de decisiones, v) evaluación de las opciones de solución, estableciendo cuál es la más adecuada de acuerdo a sus objetivos, vi) estudio de los pros y los contras de cada solución y evaluación de los resultados, vii) preparación de un plan de acción (investigación acción).**

El proyecto factible es un tipo de investigación que se define como la elaboración y desarrollo de una propuesta de un modelo operativo viable para solucionar problemas, requerimientos o nece-

sidades de organizaciones o grupos sociales, y como su nombre lo indica, “tiene un propósito de utilización o ejecución inmediata de la propuesta”. La propuesta que lo define puede referirse a la formulación de políticas, programas, tecnologías, métodos o procesos, que sólo tienen sentido en el ámbito de necesidades sociales. El proyecto factible se desarrolla a través de las siguientes etapas: **i)** el diagnóstico de las necesidades (se basa en una investigación de campo o en una investigación documental), **ii)** planteamiento y fundamentación teórica de la propuesta; **iii)** el procedimiento metodológico, **iv)** las actividades y recursos necesarios para su ejecución y **v)** el análisis de viabilidad o factibilidad del proyecto (económica, política, social, entre otros) y **vi)** la posibilidad de su ejecución. De allí que, el informe final del proyecto factible se conforma con los siguientes aspectos: Introducción, contexto de la situación, planteamiento de la necesidad, objetivos y justificación del proyecto, marco referencial, metodología, diagnóstico de necesidades, formulación de la propuesta, análisis de factibilidad, recomendaciones y lista de referencias. Las etapas de un proyecto factible son: **i)** ¿Qué hacer? Se determina la identidad, naturaleza y contexto del proyecto: social, económico, político y cultural. Se busca definir y explicar los objetivos que guían la acción; **ii)** ¿Para qué hacerla? Se relaciona con la contribución a la solución de un problema o la satisfacción de una necesidad; **iii)** ¿Por qué hacerla? Se justifica el proyecto a través de una explicación sobre su importancia, viabilidad, sustento teórico, beneficiarios, interés, relevancia y motivo. También, se destacan las oportunidades actuales y a futuro que genera el proyecto; **iv)** ¿Cómo hacerlo? Se incluyen todas las actividades a realizar durante el proceso, las estrategias y tareas como parte de un plan de acción en función de los objetivos propuestos; **v)** ¿Dónde hacerlo? Se especifica el lugar donde se ejecutarán las actividades como parte del proyecto; **vi)** ¿Qué magnitud tiene? Se delimita el alcance cuantitativo del proyecto y su profundidad, el tipo de servicio que prestará o las necesidades que se propone satisfacer; **vii)** ¿Cuándo se hará? Se refiere al tiempo que se requiere para llevar a la práctica el proyecto, el cual necesariamente será objeto de un control y delimitación cronológico; **viii)** ¿Quié-

nes lo harán? Se especifican las competencias y preparación de los miembros que conforman el equipo de trabajo que pondrá en práctica el proyecto; **ix) ¿Con qué medios y recursos se hará?** Se precisan los recursos humanos, económicos, materiales y tecnológicos que se requieren para realizar el proyecto; **x) ¿Qué sucede durante la ejecución?** Se establecen los procesos de control, evaluación y seguimiento de cada una de las actividades a realizar; **xi) ¿Cuáles son las limitaciones?** Se vinculan las fortalezas, amenazas y oportunidades del proyecto, a fin de elaborar un marco de recomendaciones para su ejecución o para su réplica. Dado que un proyecto factible consiste en la investigación, elaboración y desarrollo de una propuesta, de un modelo operativo viable para solucionar problemas, requerimiento o necesidades de organizaciones o grupos sociales; puede referirse a la formulación de políticas, programas, tecnologías, métodos o procesos:”, además “es una proposición sustentada en un modelo operativo factible, orientada a resolver un problema planteado o satisfacer necesidades en una institución o campo de interés nacional”; se requiere de un esquema para su elaboración: **i) Título del Proyecto;** **ii) Introducción:** Breve reseña o explicación del proyecto, propósito del proyecto, importancia y justificación del proyecto, identificación de las partes o capítulos del proyecto; **iii) Formulación del Proyecto:** Situación Inicial: se debe señalar la situación de insatisfacción que se quiere superar o enfrentar con la ejecución del proyecto, objetivo general: describir cual es el propósito o el fin último que se desea lograr o alcanzar con la ejecución del proyecto, ubicación geográfica del proyecto: se debe precisar la localidad, o municipio, donde piensa ejecutar el proyecto que está presentando, tipo de proyecto: se debe señalar si se trata de un proyecto nuevo, una ampliación y/o consolidación de un proyecto existente, área del proyecto: se debe señalar si el área en la que se enmarca el proyecto: en el área deportiva, cultural o social – comunitaria, académica o de bienes y servicios, características generales del proyecto: se deben describir las características más destacadas del proyecto que permita conocer sus alcances, beneficiarios del proyecto: se debe señalar los beneficiarios directos (a quienes se le otorgarán los bienes y/o servicios generados

por el proyecto) y los beneficiarios indirectos (consecuencia de la ejecución del proyecto quienes se beneficiarán a futuro o en forma indirecta), tiempo de ejecución del proyecto: se debe señalar el tiempo previsto (semanas o meses) para la ejecución del proyecto, costo total del proyecto: se debe determinar el monto total estimado para la ejecución del proyecto, discriminando el monto solicitado, presentado por, fecha y firma: se escribe el nombre de la persona que presenta el proyecto, se firma y se coloca la fecha en que se consigna; **iv) Justificación, Objetivos y Metas:** Se definen las razones que justifican el proyecto, los objetivos y metas. Los objetivos expresan lo que se quiere lograr con la ejecución del mismo; es el motor que le da dirección a todas sus acciones. Justificación: Se describen los argumentos que justifican la ejecución del proyecto. Se requiere responder: ¿Cuál es la prioridad y urgencia de ejecutar el proyecto? ¿Por qué el proyecto que se plantea es el más adecuado o viable?, Objetivo General: Se identifica el propósito central del proyecto y a dónde se quiere llegar. Se describe la situación ideal futura que se desea alcanzar, a través del Proyecto (este objetivo está concebido a largo plazo). Se debe responder a la pregunta: ¿Qué se quiere lograr con la ejecución del proyecto?, objetivos específicos: Se señalan los resultados parciales que debe alcanzar el proyecto para lograr el objetivo general (están concebidos a corto plazo). Se debe responder a la pregunta: ¿Cuáles son los resultados parciales o intermedios que hay que lograr para alcanzar el o los objetivo (s) general (es) del proyecto?, Metas: Se definen en correspondencia con los objetivos específicos diseñados. Se requiere que los objetivos sean traducidos en acciones concretas, es decir, hay que indicar cuánto se quiere lograr con la realización del proyecto dentro de un plazo determinado y en un ámbito o espacio también delimitado, permitiendo establecer las actividades y acciones correspondientes; **v) Estudio Social:** Se requiere un estudio social a fin de determinar los beneficiarios del proyecto, en función de las características socio-económicas, se señalan los alcances sociales que se obtendrán y cómo participarán los beneficiarios en la ejecución y operación del proyecto, se señalan los beneficiarios del proyecto. Los Beneficiarios Directos: se identifican quienes serán los

beneficiarios inmediatos, los directamente favorecidos por la consecución de los objetivos y metas del proyecto, se señala los grupos que serán receptores directos de las acciones contempladas en el proyecto. Los Beneficiarios Indirectos: se señala quienes serán los beneficiarios finales del proyecto. Se estudian las características socio-económicas del área donde se desarrollara, es decir, se debe identificar y dar conocer las necesidades del área donde se va a desarrollar el proyecto, destacando aspectos socio - económicos y características y establezcan relaciones con diagnósticos lo cual permita tener un panorama general de las condiciones de dicha área. Se deben obtener fuente de información que expliquen de donde obtuvo los datos que caracterizan el área objeto de este proyecto. Se debe exponer el beneficio a la comunidad señalando como el proyecto cubrirá la necesidad previamente detectada y cómo la comunidad se beneficiará con los bienes y/o servicios proporcionados. Se debe incorporar la participación de la comunidad en el proyecto señalando qué tipo de estrategia se utilizará para incorporarla en la formulación y ejecución del mismo ya que ello afectará el grado de incorporación del proyecto en su sostenibilidad. Se deben evaluar los impactos esperados describiendo en términos cuantitativos y/o cualitativos las modificaciones en la situación problema descrita que se espera alcanzar con la ejecución del proyecto tomando en cuenta la población beneficiaría y su contexto. Las descripciones cuantitativas se deberán enunciar a través de indicadores que expresan cantidades, tales como números, porcentajes, frecuencias, promedio, tasas, etc. Las descripciones cualitativas expresan calidad, tal como, clase, tipo, variedad, diferenciación; **vi) Consideraciones finales para la viabilidad y factibilidad del proyecto:** Se debe redactar, en síntesis, las consideraciones del autor o los autores del proyecto, con relación a la viabilidad (el camino a seguir) y la factibilidad (posibilidad de realizar el proyecto) con el propósito de mostrar su perspectiva en los aspectos económicos, sociales e institucionales.

La investigación acción, es un tipo de investigación que pretende enlazar el enfoque experimental de la ciencia social con programas de acción social que respondan a los problemas sociales que emer-

gen de lo habitual. La investigación-acción inicia el cuestionamiento del fenómeno social con la pretensión de tratar de forma simultánea conocimientos y cambios sociales, de manera que se unan la teoría y la práctica. La investigación-acción es una forma de indagación introspectiva colectiva emprendida por participantes en situaciones sociales que tiene el objeto de mejorar la racionalidad y la justicia de sus prácticas sociales, así como su comprensión de esas prácticas y de las situaciones en que éstas tienen lugar. Ejemplo: indagación para: igualación de oportunidades para obtener empleo, la causa y curación de perjuicios en los niños, la socialización de las bandas callejeras,... El concepto tradicional de investigación-acción proviene de las tres etapas del cambio social: descongelamiento, movimiento y recongelamiento. Las etapas de la metodología de la investigación-acción son: *Insatisfacción con el actual estado de cosas, Identificación de un área problemática, Identificación de un problema específico a ser resuelto mediante la acción, Formulación de varias hipótesis, Selección de una hipótesis, Ejecución de la acción para comprobar la hipótesis, Evaluación de los efectos de la acción, Generalizaciones*. En la investigación – acción, el quehacer científico consiste no solo en la comprensión de los aspectos de la realidad existente, sino también en la identificación de las fuerzas sociales y las relaciones que están detrás de la experiencia humana. Su función fundamental es servir a la acción. La investigación acción tiene como perspectiva la transformación de la realidad a través de la práctica, es decir, la planificación, pasa a la acción, para luego a la observación y finalmente con una determinada reflexión.

La investigación y desarrollo (I+D), se caracteriza porque realiza conjunto de actividades emprendidas de forma sistemática, a fin de aumentar el caudal de conocimientos científicos y técnicos, así como la utilización de los resultados de estos trabajos para conseguir nuevos dispositivos, productos, materiales o procesos. Es un tipo de investigación que se aplica a los departamentos de investigación públicos o privados encaminados al desarrollo de nuevos productos o la mejora de los existentes por medio de la investigación científica.

Y **la investigación, desarrollo e innovación (I + D + i)**, es un tipo de investigación que involucra tres aspectos: **investigación (i)**, que es la indagación original planificada que persigue descubrir nuevos conocimientos en el ámbito científico y tecnológico, **desarrollo (D)**, que es la aplicación de los resultados de la investigación para la fabricación de nuevos productos y para el diseño de nuevos procesos en sistemas de producción, así como para la mejora tecnológica de los procesos o sistemas preexistentes, e **innovación tecnológica (i)**, que es la actividad cuyo resultado sea un avance e innovación tecnológica en la obtención de nuevos productos y procesos de producción mejorando sustancialmente los ya existentes.

También se pueden mencionar otros tipos de investigación como la **investigación participativa** que se define como forma de investigación que es considerada como un proceso sistemático que lleva a cabo una determinada comunidad para llegar a un conocimiento más profundo de sus problemas y tratar de solucionarlos, intentando implicar a toda la comunidad en el proceso. Este tipo de investigación determina una problemática a ser resuelta dentro de una comunidad pequeña, ya sea barrio, pueblo o comarca; involucra a la comunidad en el proyecto con la finalidad de producir un resultado donde la misma salga favorecida, de allí conjuntamente instituir reflexiones críticas a la acción transformadora tomando el diálogo como una técnica de igualdad que más adelante produzca experiencias significativas.

La investigación colaborativa que se define como forma de investigación en que varias entidades – generalmente de diferente naturaleza – comparten el interés por la ejecución de un proyecto, el esfuerzo por desarrollarlo, los riesgos y la propiedad de los resultados conforme a su diversa contribución para obtenerlos. Los participantes en un proyecto colaborativo se asocian según las condiciones pactadas en un acuerdo de consorcio que regula los procedimientos de toma de decisiones y estructura del consorcio, la contribución y papel específico de cada socio, y la propiedad de los resultados, entre otras cuestiones. Los aspectos clave que deben negociarse con

mayor detenimiento suelen afectar a la propiedad del conocimiento previo que se aporta por cada socio, a la propiedad de los resultados que surjan del proyecto, y los cauces que se prevé ayudarán a la explotación de los mismos.

La **investigación social** que se define como forma de investigación que utilizando el método científico, permite obtener nuevos conocimientos en el campo de la realidad social (investigación pura) o que permite estudiar una situación social para diagnosticar necesidades y problemas a los efectos de aplicar los conocimientos con finalidades prácticas (investigación aplicada) y la **investigación educativa** es un proceso planificado de *acción, observación, reflexión y evaluación de forma cíclica* que se propone intervenir en la práctica educativa para su mejoramiento implicando en eso a los actores sociales, los cuales conducen y negocian la innovación educativa. En este tipo de investigación el investigador está inmerso en una espiral auto-reflexiva, porque este proceso es un “*ir y venir*” entre la “*acción y la reflexión*” (*espiral dialéctica*); es aquí donde se alternan e integran momentos del proceso investigativo para generar la transformación nacida del mismo seno de los actores sociales y el investigador, por lo cual esta investigación es de tipo idiográfica.

Para finalizar el capítulo se hace énfasis en que todo proceso de investigación en ingeniería debe estar enfocado a dar *solución de problemas sociales*, lo que consiste en dar respuesta a una necesidad colectiva. Los problemas, son asuntos que implican algún tipo de inconveniente o trastorno y que exigen una solución. Cuando aparece un problema, éste supone una dificultad para alcanzar un objetivo. Lo social, por su parte, es un adjetivo que refiere a lo que está vinculado a la sociedad. Una sociedad es una comunidad formada por personas que interactúan entre sí y que tienen una cultura en común, dado que “*sociedad*” es un término que describe a un grupo de individuos marcados por una cultura en común, un cierto folclore y criterios compartidos que condicionan sus costumbres y estilo de vida y que se relacionan entre sí en el marco de una comunidad”. El desarrollo científico universitario con miras a resolver problemas so-

ciales presenta múltiples perspectivas desde las cuales se debe encarar el problema de la relación entre la universidad, la investigación y la ciencia. De ahí, la importancia de que la investigación y el desarrollo científico universitario tengan miras a resolver problemas sociales, económicos y educativos de una región en la que la Universidad está inserta y concentrarse en las estrategias para sacar provecho de la calidad académica de sus docentes y alumnos con refinados métodos e instrumentos de investigación de vanguardia. Los problemas sociales, por lo tanto, son situaciones que impiden el desarrollo o el progreso de una comunidad o de un sector de ellas. Puede decirse que un problema social surge cuando las personas de una comunidad no logran satisfacer sus necesidades básicas. Un problema social se presenta cuando un sector de la población o una comunidad no logra acceder: a los servicios de salud, de educación, al servicio agua potable, al sistema de enclocado, al servicio de luz, al sistema de distribución de alimentos, a la posibilidad de adquisición de una vivienda digna, a un nivel de seguridad y respeto por su integridad física. Los gobiernos son los encargados de desarrollar las políticas sociales necesarias que permitan revertir el problema social en cuestión y, de este modo, mejorar la calidad de vida de la gente. La sociedad civil, de todos modos, también puede trabajar para la solución de los problemas sociales. Existen numerosas organizaciones no gubernamentales que tienen como finalidad combatir los problemas sociales a través de campañas, programas de desarrollo, etc. Aunque no existen países sin problemas sociales, la menor cantidad de ellos en la realidad nacional supone un indicador de desarrollo. A menor incidencia de los problemas sociales en la vida de la población, mayor progreso. Para resolver un problema social hay que: **i) identificar las situaciones problemáticas, ii) describir con detalle el problema y la respuesta habitual a dicho problema. Al describir la situación y la respuesta en términos de quién, qué, dónde, cuándo, cómo y porqué, se verá el problema de forma más clara, iii) hacer una lista con las alternativas. En esta fase se utiliza la estrategia denominada “tempestad de ideas” para conseguir los objetivos recientemente formulados. La**

técnica de la tempestad de ideas debería limitarse, durante esta fase, a estrategias generales para alcanzar los objetivos, **iv)** ver las consecuencias. Este paso consiste en seleccionar las estrategias más prometedoras y evaluar las consecuencias de ponerlas en práctica y **v)** evaluar los resultados: Una vez se haya intentado la respuesta nueva, se deberán observar las consecuencias, por ejemplo: ¿Suceden las cosas tal como estaban previstas? Si no es así se buscan nuevas alternativas al problema. Como ilustración de problemas sociales se pueden mencionar: **i)** Como una comunidad puede generar un mecanismo de reciclaje de la basura en donde el reciclaje de ciertos rubros generen dividendos al colectivo, **ii)** Como diseñar un sitio en la Web a través de una plataforma virtual como espacio de denuncia comunitaria de los problemas más relevantes de una comunidad para ser llevados en consenso a los entes pertinentes para su pronta atención, **iii)** Como informar a una comunidad respecto a las operativos de vacunación y asistencia médica a través de las nuevas tecnologías de la información y la comunicación, **iv)** Como crear un red local de oferta y demanda de servicios pagados para resolver problemas específicos en la zona.

La solución de un problema social, es un proceso consciente, racional, que requiere esfuerzo y de naturaleza cognitivo-conductual por el que un individuo o grupo de individuos tratan de identificar soluciones efectivas a problemas específicos que aparecen en la vida cotidiana. Un proceso investigativo, permite encontrar una variedad de soluciones efectivas e incrementa la probabilidad de seleccionar la más eficaz. Para este propósito es necesaria la realización previa de un “análisis sistémico” que se puede definir como una estructura teórica o mapa conceptual de relaciones y ordenes que relacione, unifique, integre y organice los elementos aparentemente aislados de un fenómeno social en una nueva totalidad mediante la dinámica de sus componentes y las relaciones entre ellos. El método sistémico estudia los elementos interactivos de los fenómenos desde múltiples puntos de vista con la intencionalidad de integrar, en lo posible, esas perspectivas, ofreciendo panoramas coherentes y estructurado del estado general. Por ejemplo, si en el país se tiene un problema edu-

cacional por el cual existe demasiada deserción escolar. El método sistémico comienza a operar por identificar el sistema en el que se encuentra inmersa la deserción escolar. Descubre que dicho sistema tiene componentes como el profesor, el colegio, los compañeros de clase, los cursos y los planes de estudio. Al continuar apreciando en realidad, establece que el profesor, su tecnología educativa y el contenido de los cursos no son nada atractivos y que provocan el rechazo del alumno; por lo tanto, para erradicar la deserción se deben modificar drásticamente estos sub-sistemas y componentes de la educación.

El mecanismo o vaso comunicante digital para el planteamiento y solución de problemas sociales actualmente entre “*universidad, ingeniero de sistemas y sociedad*” son las TICs, que representan el conjunto de tecnologías de la información y comunicación soportadas en las herramientas telemáticas desarrolladas para gestionar información y enviarla de un lugar a otro y que funcionan como interfaz de recepción, almacenamiento, procesamiento y envío de información a través de una gran gama de dispositivos electrónicos en forma digitalizada. A través de las TICs y de sus dispositivos electrónicos fijos y portátiles es posible crear espacios digitales de comunicación a distancia para: **i) canalizar y dar asesoría a problemas comunales, ii) crear puentes de enseñanza/ aprendizaje comunal colaborativa, y iii) plantear investigaciones que resuelvan problemas comunales y/o sociales.**

Actividades del capítulo II

1. ¿Cuál debe ser la metodología investigativa aplicada por el ingeniero de sistemas al realizar un estudio o una investigación?
2. ¿En qué consiste una investigación científica y qué se debe definir o determinar para comenzar a realizar una investigación científica?
3. ¿Cuál es la importancia de la aplicación del método científico a la hora de realizar una investigación en ingeniería de sistemas y cuáles son sus etapas del método científico y en que radica su importancia?
4. ¿En qué se diferencia una ley de una teoría?
5. Explique en qué consiste el diseño y contextualización de una investigación.
6. Explique en qué consiste una investigación positivista, fenomenológica y psicoanalítica.
7. Explique en qué consiste una investigación de carácter exploratorio, de carácter descriptivo y de carácter explicativo.
8. Explique en qué consiste el paradigma cuantitativo, el paradigma cualitativo y el paradigma holístico.
9. ¿En qué consiste metodología de la investigación cuantitativa?
10. Explique en qué consiste la investigación experimental y la investigación no experimental.
11. Explique en qué consisten los experimentos puros, los cuasi-experimentos, la investigación de campo, la investigación descriptiva y sus ramificaciones.

12. Explique en qué consiste la investigación pre experimental, la investigación ex post facto, la investigación transeccional o transversal y sus ramificaciones, la investigación longitudinal y sus ramificaciones.
13. ¿En qué consiste metodología de la investigación cualitativa?
14. Explique en qué consiste la investigación hermenéutica, la investigación etnográfica, la investigación etnometodológica, la investigación documental y la investigación biográfica.
15. ¿En qué consiste metodología de la investigación holística?
16. Explique en qué consiste un proyecto factible, la investigación acción, la investigación y desarrollo (I+D), la investigación, desarrollo e innovación (I + D + i).
17. Explique en qué consiste un análisis sistémico.
18. Explique con sus palabras cómo integran las TIC's para el planteamiento y solución de problemas sociales en donde se involucre a la Universidad, el Ingeniero y la Sociedad.

Análisis estadístico aplicado a la investigación

3.1. Metodología estadística aplicada a la investigación

El análisis estadístico de una investigación en ingeniería comienza con la estrategia para diseñar una “*metodología estadística*” o una colección de “*métodos estadísticos*” para ser aplicados al estudio que se quiere llevar cabo, lo cual va a estar determinada fundamentalmente por el “*diseño de la investigación a realizar*” en función de los objetivos propuestos o que se persigan con la misma. Un análisis estadístico comienza primeramente en el diseño de un mecanismo de muestreo para la toma una “*muestra física*” de una “*población física en estudio*”, luego, se extraen de manera los “*datos estadísticos*” de la muestra física a través de la observación o experimentación con la misma.

El análisis exhaustivo de los datos obtenidos y las conclusiones derivadas de estos, representan el “*centro de gravedad de una investigación*” porque la realización de un “*análisis descriptivo y/o inferencial*” va a permitir dar “*validez a los datos y resultados*” obtenidos de la “*observación de un fenómeno*” o de la “*experimentación con el fenómeno*”. Por un lado, el “*análisis descriptivo*” se define como el conjunto de métodos estadísticos que describen una muestra de observaciones obtenidas de un colectivo de elementos. Su objetivo fundamental es describir las diversas características observables o medibles de la población física en forma “*conveniente, útil, práctica y comprensible*” mediante el uso de “*tablas de frecuencia, gráficos y estadísticas y/o estimadores*”. Por otro lado, el “*análisis inferencial*”

se define como el conjunto de métodos que “reduce el estudio de un fenómeno a una de sus partes o a una muestra” al generalizar los resultados obtenidos de la muestra a toda la población objeto a estudio o de observación, referidos a una característica cuantificable o medible.

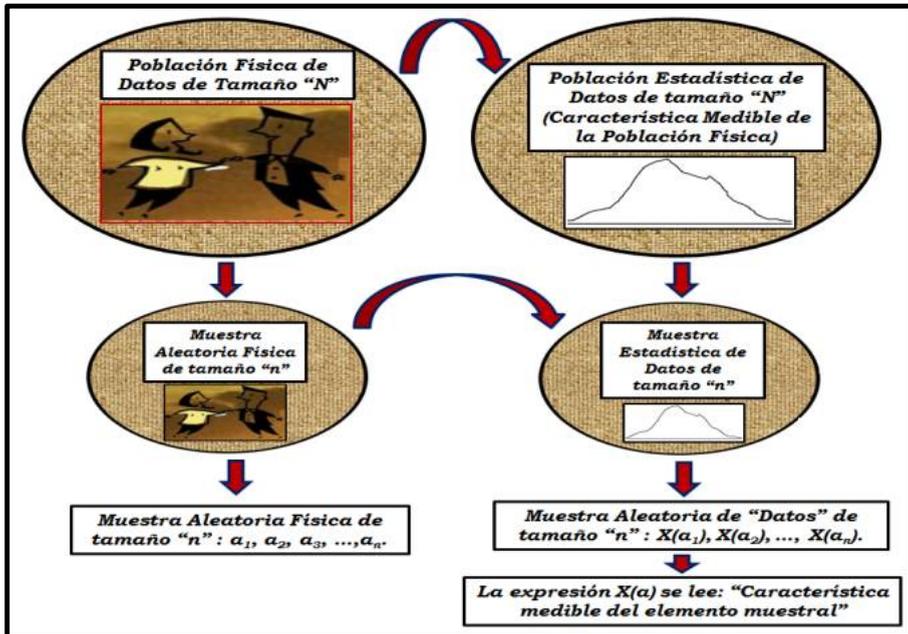


Figura 7. Metodología estadística aplicada a la investigación. Fuente: Autores.

Ejemplo 4: Una investigación durante un mes, luego del receso vacacional, en los jóvenes de una localidad suramericana referente a “la Anorexia”, ha determinado que la delgadez representa el “físico ideal” para ellos, dado que es un sinónimo de éxito social, por tanto luchan para conseguirlo motivados por los patrones impuestos por modelos, artistas y por la publicidad comercial. En esta investigación, se observó una muestra de 27 jóvenes adolescentes con síntomas de anorexia, registrándose los siguientes signos visibles:

El análisis comienza por resumir la información en una tabla, llámese “Tabla de frecuencias” de los signos visibles de 27 adolescentes con síntomas de anorexia (Tabla 10).

Tabla 10. Muestra física de 27 jóvenes adolescentes con síntomas de anorexia

Clases	Signos Visibles	Conteo	Frecuencia	Porcentaje
Clase 1	Dieta severa	////////	9	33,33%
Clase 2	Miedo a engordar	///	3	11,11%
Clase 3	Hiperactividad	////	4	14,81%
Clase 4	Uso de laxantes	////	5	18,52%
Clase 5	Uso de ropa holgada	//////	6	22,22%
	Total		27	100%

Fuente: Autores.

Gráficamente se puede observar en las Figuras 8 y 9.

El análisis realizado, permitió organizar la información de una forma “útil, práctica y conveniente” a través de una tabla de frecuencias, un gráfico de barras y un gráfico circular que ilustra de forma porcentual los “signos visibles de la anorexia” predominantes por tener una mayor frecuencia y se expresan en forma porcentual observándose que: la “dieta severa” tiene un 33%, seguido por “uso de ropa holgada” que tiene un 22% y “uso de laxantes” que tiene un 19%.

Ejemplo 5: Una investigación durante una semana, respecto a las “preferencias electorales” de los candidatos A, B, C y D, de una “muestra aleatoria (al azar)” de un estrato de “45 jóvenes de sexo femenino” entre edades de 18 años a 25 años en una provincia de un país de Centroamérica “población en estudio”, arrojó los datos que se registran en la Tabla 12.

El análisis comienza por un resumen descriptivo de la información a través de una Tabla de frecuencias, de las preferencias electorales del estrato de las “45 jóvenes de sexo femenino” entre edades de 18 años a 25 años (Tabla 13).

Tabla 11. Signos visibles de 27 jóvenes adolescentes con síntomas de anorexia.

Muestra Física de Adolescentes con Anorexia (n = 27)					
Muestra Física (del 1 al 9)	Características observable	Muestra Física (del 10 al 18)	Características observable	Muestra Física (del 19 al 27)	Características observable
Adolecente 1	Dieta severa	Adolecente 10	Miedo a engordar	Adolecente 19	Hiperactividad
Adolecente 2	Uso de ropa holgada	Adolecente 11	Dieta severa	Adolecente 20	Uso de laxantes
Adolecente 3	Miedo a engordar	Adolecente 12	Dieta severa	Adolecente 21	Uso de ropa holgada
Adolecente 4	Dieta severa	Adolecente 13	Uso de ropa holgada	Adolecente 22	Dieta severa
Adolecente 5	Dieta severa	Adolecente 14	Dieta severa	Adolecente 23	Uso de ropa holgada
Adolecente 6	Hiperactividad	Adolecente 15	Uso de laxantes	Adolecente 24	Miedo a engordar
Adolecente 7	Uso de laxantes	Adolecente 16	Dieta severa	Adolecente 25	Uso de ropa holgada
Adolecente 8	Uso de laxantes	Adolecente 17	Hiperactividad	Adolecente 26	Uso de laxantes
Adolecente 9	Uso de ropa holgada	Adolecente 18	Hiperactividad	Adolecente 27	Dieta severa

Fuente: Autores.

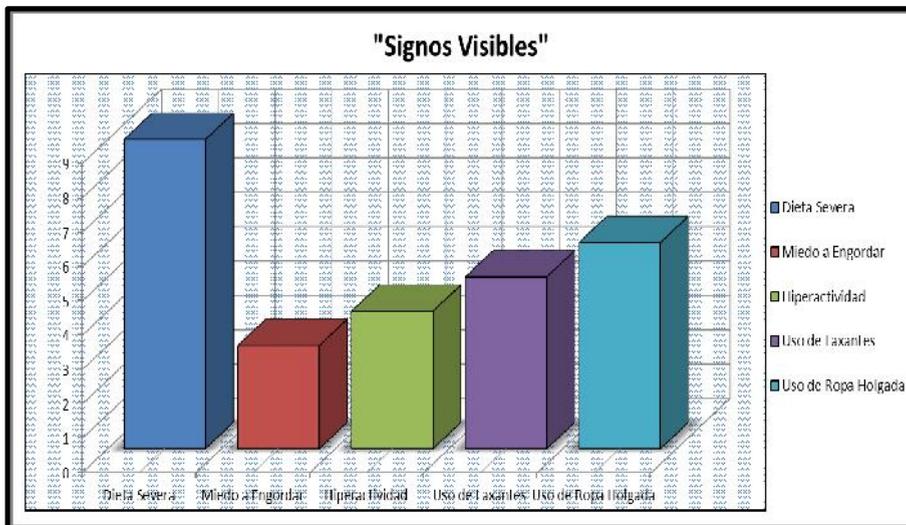


Figura 8. Signos visibles de 27 jóvenes adolescentes con síntomas de anorexia barra. Fuente: Autores.

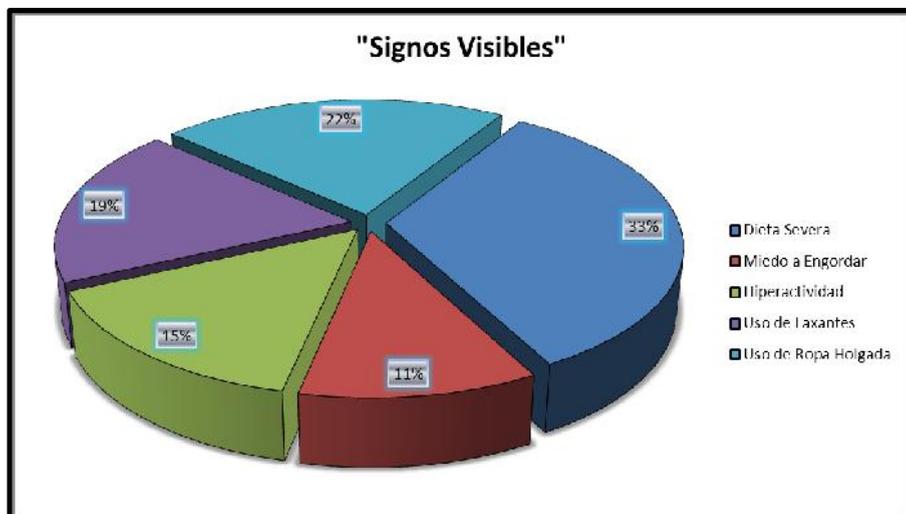


Figura 9. Signos visibles de 27 Jóvenes adolescentes con síntomas de anorexia torta. Fuente: Autores.

Tabla 12. Muestra aleatoria de un estrato de 45 jóvenes de sexo femenino

Muestra física de 45 jóvenes de sexo femenino (n = 45)					
Muestra Física (del 1 al 15)	Característica Observable	Muestra Física (del 16 al 30)	Característica Observable	Muestra Física (del 31 al 45)	Característica Observable
Joven 1	Candidato A	Joven 16	Candidato C	Joven 31	Ninguno
Joven 2	Candidato A	Joven 17	Candidato A	Joven 32	Candidato C
Joven 3	Candidato B	Joven 18	Ninguno	Joven 33	Ninguno
Joven 4	Candidato D	Joven 19	Candidato B	Joven 34	Candidato A
Joven 5	Candidato B	Joven 20	Candidato D	Joven 35	Candidato A
Joven 6	Candidato C	Joven 21	Candidato B	Joven 36	Ninguno
Joven 7	Candidato C	Joven 22	Candidato A	Joven 37	Candidato A
Joven 8	Candidato A	Joven 23	Candidato C	Joven 38	Candidato B
Joven 9	Ninguno	Joven 24	Ninguno	Joven 39	Candidato A
Joven 10	Candidato D	Joven 25	Candidato C	Joven 40	Candidato B
Joven 11	Candidato D	Joven 26	Candidato A	Joven 41	Candidato A
Joven 12	Candidato A	Joven 27	Candidato A	Joven 42	Candidato C
Joven 13	Candidato B	Joven 28	Ninguno	Joven 43	Candidato A
Joven 14	Candidato B	Joven 29	Candidato B	Joven 44	Candidato B
Joven 15	Ninguno	Joven 30	Ninguno	Joven 45	Candidato D

Fuente: Autores.

Tabla 13. Preferencias electorales de un estrato de 45 jóvenes de sexo femenino

Clases	Preferencias Electorales	Frecuencia	Porcentaje	Intervalo de 90% de Confiabilidad
Clase 1	Candidato A	14	31,11%	Desde 19,76% hasta 42,46%
Clase 2	Candidato B	10	22,22%	Desde 12,03% hasta 32,41%
Clase 3	Candidato C	7	15,56%	Desde 6,67% hasta 24,45%
Clase 4	Candidato D	5	11,11%	Desde 3,40% hasta 18,82%
Clase 5	Ninguno	9	20,00%	Desde 10,19% hasta 29,81%
	Total	45	100%	

Fuente: Autores.

Gráficamente se puede observar:

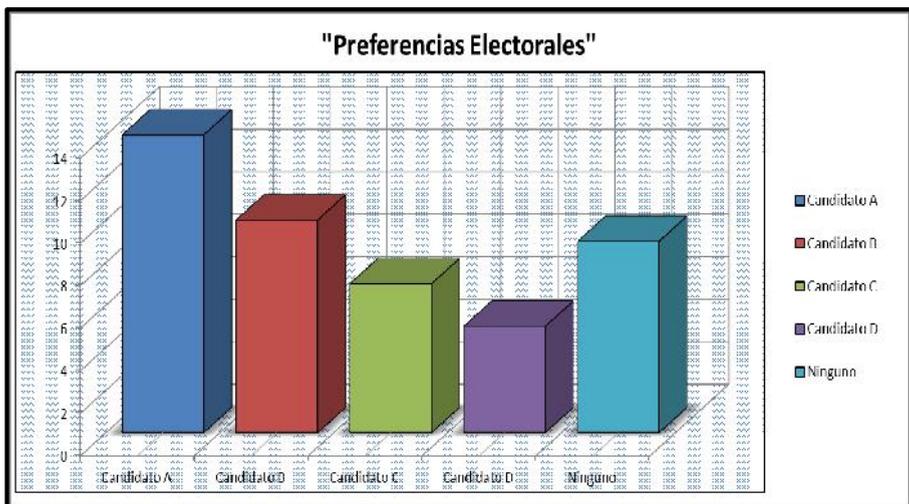


Figura 10. Preferencias electorales de un estrato de 45 jóvenes de sexo femenino.
Fuente: Autores.

El proceso estadístico va a consistir en generalizar los resultados obtenidos de la “muestra aleatoria” a toda la población objeto de estudio “Provincia de un país de Centroamérica”. La tabla de frecuencia ilustra los porcentajes de las preferencias obtenidas en la muestra. El proceso inferencial, va a consistir en generalizar estos

resultados a toda la población en estudio, este caso representado por la columna de “Intervalos de 90% de confianza” (el mecanismo de realización de estos intervalos se verá en capítulos posteriores). De ahí que, al realizar un proceso inferencial se tenga que: el “Candidato A” tiene un porcentaje de votos en la provincia entre un 19,76% y un 42,46%, el “Candidato B” tiene un porcentaje de votos en la provincia entre un 12,03% y un 32,41%, el “Candidato C” tiene un porcentaje de votos en la provincia entre un 6,67% y un 24,45%, el “Candidato D” tiene un porcentaje de votos en la provincia entre un 3,4% y un 18,82%, y “Ninguno” tiene un porcentaje de votos en la provincia entre un 10,19% y un 29,81%. Para que el margen de error del intervalo (el intervalo de porcentajes sea más pequeño), se debe tomar una muestra de tamaño más grande.

3.2. Fases de un análisis estadístico aplicado a la investigación

Las fases o etapas de este análisis estadístico de una investigación realizada en ingeniería, se pueden enumerar de la siguiente manera: *Paso 1* “Planteamiento del problema y población física a ser estudiada”, *Paso 2* “Diseño del plan de muestreo”, *Paso 3* “Diseño del mecanismo de obtención de la información”, *Paso 4* “Procesamiento y análisis de los datos”, *Paso 5* “Conclusiones resultados de la investigación”.

Paso 1) *Planteamiento del problema y población física a ser estudiada*: Consiste en describir de manera amplia y delimitar “el problema y/o la situación objeto de estudio”, ubicándola en un contexto que permita concretar “la delimitación de la población física involucrada, la delimitación de tiempo involucrado y la delimitación del espacio o lugar donde ocurre situación objeto de estudio”. Se debe plantear el problema en detalle especificando de manera exacta: la población física asociada al problema o fenómeno planteado, la pregunta o interrogante a responder, la hipótesis que responde la pregunta o interrogante si es que la hubiera y las “características o variables” que pueden ser “observables o medibles” en la población física a ser estudiada.

Paso 2) *Diseño del plan de muestreo*: De acuerdo a la “*forma de la población física*” y de cómo se definan las “*características o variables medibles*” de la investigación, se realiza el diseño del mecanismo de muestreo “*aleatorio o no aleatorio*” a utilizar para obtener la “*muestra física de elementos poblacionales*”. Se debe determinar de qué tamaño “*n*” se ha de tomar la muestra física de elementos, ya sea en función del tamaño “*N*” de la población física de elementos en estudio, o en función del nivel de confiabilidad que se requiere para estimar un determinado parámetro de una característica o variable poblacional.

Paso 3) *Diseño del mecanismo de obtención de la información*: Consiste en el esquema del mecanismo a utilizar “*experimental o no experimental*” para obtener la “*muestra de datos estadísticos*” de la “*muestra física de elementos poblacionales*”. Aquí se definen las herramientas de recolección de la información, de la muestra física de elementos, tales como: *experimentos, entrevistas, test, encuestas, diarios de campo, grabaciones, videos, documentos, mediciones sucesivas (panel), estudios de caso,...* y qué medidas se han de tomar para asegurar que estas “*observaciones o datos*” sean objetivos y precisos.

Paso 4) *Procesamiento y análisis de los datos*: Explica a través de qué mecanismos estadísticos se va a “*organizar y representar*” la “*muestra estadística*”, es decir, “*la muestra de datos numéricos o literales*” (estadística descriptiva). Es decir, se pueden organizar los datos a través de: *tablas simples, tablas de frecuencias, tablas de contingencia, matrices de datos,...* y se puede graficar los datos a través de *gráficos: circulares, de puntos, de barra, de línea, de área, de comparación, de correlación, de caja,...* y posteriormente, a través de que procedimientos o métodos estadísticos serán analizados los datos de la muestra para extraer las conclusiones del estudio, esto puede ser a través: *cálculo de estadísticas o estimadores, estimación de parámetros poblacionales, comparación de parámetros de dos o más poblaciones, etc.* Luego, de acuerdo al tipo de análisis estadístico realizado, el cual puede ser “*descriptivo*”, el cual de acuerdo al planteamiento de una hipótesis inicial va a realizar una “*descripción de la estructura*” de una situación poblacional, y/o la “*comparación de una*

o más características o variables” de la estructura de uno o más fenómenos o situaciones poblaciones estudiadas; o puede ser también de tipo “inferencial”, el cual pretende generalizar los resultados obtenidos en la “muestra estadística” a toda la “población estadística” en estudio. Aquí se fija el nivel de confiabilidad o nivel de significancia, en términos de porcentaje para la inferencia o generalización de los resultados. Se debe establecer “que” se ha encontrado informando los hallazgos obtenidos, mostrándose e ilustrándose los resultados importantes obtenidos en tablas y gráficos (Análisis cuantitativo). El proceso descriptivo o inferencial ha de conducir al analista a una o varias conclusiones que servirán de orientación para tomar una decisión acertada en la respuesta las interrogantes que se plantean en el fenómeno objeto de estudio en una determinada población física.

Paso 5) *Conclusiones de la investigación*: Se presentan en un resumen escrito de los resultados (*no se permiten gráficas o tablas*), y es la oportunidad de sacar las conclusiones finales de la investigación planteada y de hacer los últimos comentarios sobre el estudio (*Análisis cualitativo*). Esta sección también incluirá motivos por los que los resultados son relevantes, explicándose las consecuencias del análisis estadístico de los datos de la muestra, los alcances y limitaciones del análisis y deja abiertas las posibilidades para realizar nuevas investigaciones.

“Ejemplo 6”: Plantee teóricamente la formulación de un “análisis estadístico” en el que se calcule una medida que represente la calificación promedio de una prueba de “habilidad numérica” aplicada a los estudiantes de *ingeniería mecánica, electrónica, industrial, producción animal, civil y en informática*; del primer semestre de la *Universidad Simón Bolívar* en donde también se realice una “comparación del nivel de destreza numérica” encontrado en los estudiantes de las carreras de ingeniería mencionadas.

Paso 1) Se quiere calcular un “valor numérico”, llámese “estimator poblacional”, que represente el puntaje promedio obtenido por los estudiantes de cada una de las carreras ingeniería de la facultad de ingeniería de la *Universidad Simón Bolívar*, llámese “población en

estudio”, en una prueba de “habilidad numérica” en donde se pueda realizar una comparación del nivel de “destreza numérica” por carrera, llámese “característica poblacional en estudio”, sabiéndose que los estudiantes de ingeniería están estratificados por: *estudiantes de ingeniería mecánica, estudiantes de ingeniería electrónica, estudiantes de ingeniería industrial, estudiantes de ingeniería en producción animal, estudiantes de ingeniería civil y estudiantes de ingeniería en informática*. Y además, se sabe de antemano que los estudiantes de cada una de las ingenierías fueron seleccionados por tres mecanismos: admisión a través de un examen de admisión realizado por la Universidad, admisión a través de asignaciones que otorga la prueba que realiza el estado y admisión a través del rendimiento obtenido en el curso pre-universitario dictado por la Universidad. También, se plantea realizar una “comparación del nivel de habilidad numérica” encontrado en los estudiantes de cada una de las carreras de ingeniería mencionados.

Paso 2) Para tomar la muestra a la que se le va a aplicar la “prueba de habilidad numérica” primeramente se estratifica o se procede a particionar la población universitaria por carrera, determinándose el número de estudiantes inscritos por cada una. Después, se calcula el porcentaje de estudiantes inscritos por cada carrera. Luego, se procede a tomar una muestra proporcional al número de estudiantes por estrato o carrera y en función del tamaño de la población física de estudiantes de ingeniería, llámese “N”. Supóngase que el número de estudiantes de ingeniería es de 7000 ($N = 7000$) y el número de estudiantes por carrera está representado por la “tabla de porcentajes” (Tabla 14) y por la Figura 11.

Una vez determinado el tamaño de la muestra “n” a tomar, la forma de ésta podría ser “estratificada o particionada” por carrera, tomada con un criterio en donde el número de elementos de la muestra sean tomados “al azar o por sorteo” realizando el procedimiento en cada estrato o carrera, y en donde el tamaño sub-muestral tomado, sea proporcional al porcentaje de elementos contenidos en cada estrato poblacional. De esta forma, la muestra física de estudiantes

Tabla 14. Comparación del nivel de destreza numérica por carrera (1)

Población física de tamaño “N = 7000” estudiantes de ingeniería	Número de estudiantes por carrera	Porcentaje de estudiantes por carrera
Ingeniería Mecánica	950	13,57%
Ingeniería Electrónica	770	11,00%
Ingeniería Industrial	1600	22,86%
Ingeniería en Producción Animal	900	12,86%
Ingeniería Civil	1100	15,71%
Ingeniería Ambiental	480	6,86%
Ingeniería Informática	1200	17,14%
Total	7000	100%

Fuente: Autores.

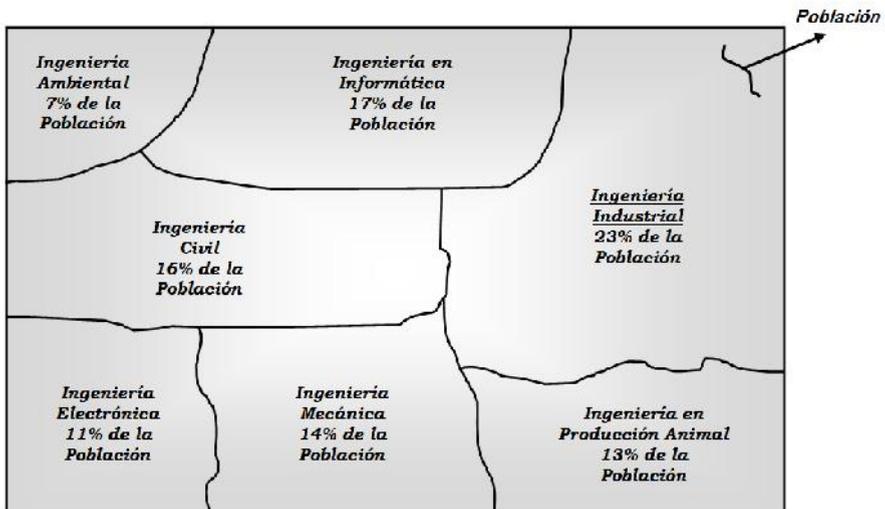


Figura 11. Estratificación de la población de estudiantes de ingeniería de la U.S.B. autores. Fuente: Autores.

que presentará la prueba deberá estar compuesta aproximadamente por un: 14% de los estudiantes tomados al azar de *ingeniería mecánica*, 11% de los estudiantes tomados al azar de *ingeniería electrónica*, 23% de los estudiantes tomados al azar de *ingeniería industrial*, 13% de los estudiantes tomados al azar de *ingeniería en producción animal*, 16% de los estudiantes tomados al azar de *ingeniería civil*, 7% de los estudiantes tomados al azar de *ingeniería ambiental* y 17% de los estudiantes tomados al azar de *ingeniería informática*.

Paso 3) El diseño del mecanismo a utilizar para obtener la “*muestra de datos estadísticos*” de la “*muestra física estratificada de estudiantes*” tomada por carrera de ingeniería, se hará a través de un “*test o examen*” de habilidad numérica. El conjunto de los resultados obtenidos por cada estudiante de cada carrera, representara la “*muestra estadística*” de datos numéricos a ser analizados y sometidos a las pruebas estadísticas pertinentes o escogidas en función de los objetivos perseguidos. Cada estudiante de la muestra física proveerá un dato numérico a la muestra estadística. Supóngase que los “*puntajes o datos*” obtenidos por los estudiantes de ingeniería, se ilustran en la Tabla 15.

Es de notar que el tamaño de la muestra es:

$$n = n_1 + n_2 + n_3 + n_4 + n_5 + n_6 + n_7$$

Paso 4) Los “*puntajes o datos*” obtenidos en la prueba de “*habilidad numérica*” por los estudiantes de la muestra estadística conformada por los estudiantes escogidos aleatoriamente de cada una de las ingenierías, como consecuencia de realizar el Paso 3), serán primeramente organizados a través de una tabla para organizar los “*resultados de la prueba por cada carrera*”, en donde se ilustra el puntaje promedio por carrera y el porcentaje de estudiantes por carrera (*en la muestra*). Supóngase que los promedios son los mostrados en la “*tabla de porcentajes*” (Tabla 16).

Tabla 15. Comparación del nivel de destreza numérica por carrera (2)

Puntajes en Ingeniería Mecánica	Puntajes en Ingeniería Electrónica	Puntajes en Ingeniería Industrial	Puntajes en Ingeniería Producción A.	Puntajes en Ingeniería Civil	Puntajes en Ingeniería Ambiental	Puntajes en Ingeniería Informática
a_1	b_1	c_1	d_1	e_1	f_1	g_1
a_2	b_2	c_2	d_2	e_2	f_2	g_2
a_3	b_3	c_3	d_3	e_3	f_3	g_3
.
.
.
a_{n1}	b_{n2}	c_{n3}	d_{n4}	e_{n5}	f_{n6}	g_{n7}

Fuente: Autores.

Tabla 16. Comparación del nivel de destreza numérica por carrera (4)

Estratos o carrera	Promedio obtenido en la prueba sobre 100	Porcentaje de estudiantes por carrera
Ingeniería Mecánica	75	14%
Ingeniería Electrónica	77	11%
Ingeniería Industrial	70	23%
Ingeniería en Producción Animal	55	13%
Ingeniería Civil	80	16%
Ingeniería Ambiental	60	7%
Ingeniería Informática	65	17%

Fuente: Autores.

De acuerdo con la Tabla 16, se puede determinar que los estratos más grandes que presentaron la prueba fueron los de *ingeniería industrial*, *ingeniería informática*, *ingeniería civil* e *ingeniería mecánica*, y los estratos más pequeños que presentaron la prueba fueron los de *ingeniería ambiental*, *ingeniería electrónica* e *ingeniería en producción animal*. También se puede determinar que los estratos que obtuvieron el mejor rendimiento en la prueba, fueron los de *ingeniería civil*, *ingeniería electrónica*, *ingeniería mecánica* e *ingeniería industrial*, y los estratos que no obtuvieron el mejor rendimiento en la prueba, fueron los de *ingeniería en producción animal*, *ingeniería ambiental* e *ingeniería informática*.

Para graficar, se puede utilizar un gráfico de barras y un gráfico circular, que permita diferenciar los rendimientos promedios obtenidos por cada carrera en la prueba (Figuras 12 y 13).

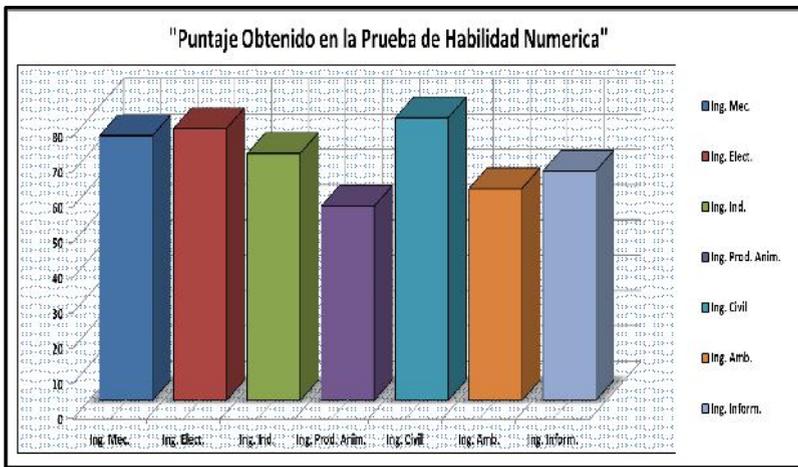


Figura 12. Comparación del nivel de destreza numérica por carrera (1). Fuente: Autores.

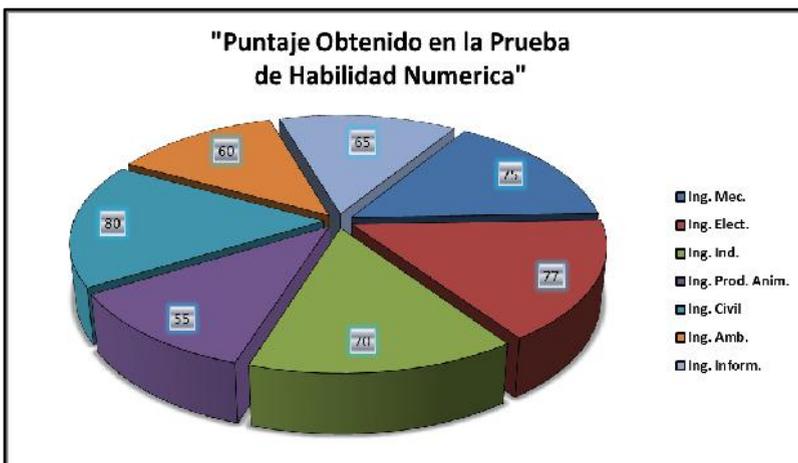


Figura 13. Comparación del nivel de destreza numérica por carrera (2). Fuente: Autores.

Además, se puede calcular un “promedio ponderado” (\bar{x}_p), llámese “estimador poblacional del promedio” de todas las carreras; es decir, un promedio que toma en cuenta el porcentaje de estudiantes que presentaron la prueba por cada “estrato o carrera” para calcu-

lar un estimador poblacional estadístico, del promedio global de la prueba de “habilidad numérica” presentada.

$$\bar{X}_p = \frac{\sum_{i=1}^7 x_i w_i \%}{\sum_{i=1}^7 w_i \%} = \frac{75 \times 14\% + 77 \times 11\% + 70 \times 23\% + 55 \times 13\% + 80 \times 16\% + 60 \times 7\% + 65 \times 17\%}{14\% + 11\% + 23\% + 13\% + 16\% + 7\% + 17\%}$$

$$\Rightarrow \bar{X}_p = \frac{7027\%}{101\%} = \frac{7027}{101} = 69,57 \Rightarrow \bar{X}_p \approx 70 \text{ puntos.}$$

El promedio ponderado de las calificaciones promedio por cada carrera o estrato es de 70 puntos. Nótese que estos cálculos son aproximados, de ahí que 100% se vea reflejado en la ecuación como un 101%.

Paso 5) Se concluye que la carrera de Ingeniería de la “Universidad Simón Bolívar” esta estratificada en 7 ingenierías con un porcentaje de estudiantes en donde predomina por los cupos asignados *ingeniería industrial, ingeniería informática, ingeniería civil e ingeniería mecánica*; y en donde los estratos más pequeños son los de *ingeniería ambiental, ingeniería electrónica e ingeniería en producción animal*. De acuerdo al análisis de las muestras estratificadas por carrera, las puntuaciones en la prueba de habilidad numérica más altas fueron en los estratos de *ingeniería civil, ingeniería electrónica, ingeniería mecánica e ingeniería industrial*, lo que de acuerdo a documentación revisada anteriormente, estos estratos tienen un nivel de exigencia numérica más alto en el examen de admisión y el curso pre-universitario de nivelación, y los estratos que no obtuvieron el mejor rendimiento en la prueba, fueron los de *ingeniería en producción animal, ingeniería ambiental e ingeniería informática*, lo que de acuerdo a documentación revisada anteriormente, estos estratos tienen un nivel de exigencia más bajo en la prueba de admisión y el curso pre-universitario de nivelación. El promedio ponderado de la prueba fue de 70 puntos y los estratos de *ingeniería civil, ingeniería electrónica e ingeniería mecánica* tienen promedios por encima de este puntaje; y los estratos de *ingeniería producción animal, ingeniería informática e ingeniería ambiental* tienen promedios por debajo de este puntaje; por otro

lado, el estrato *ingeniería industrial* coincide con este puntaje. La importancia de la determinación de los niveles de habilidad numérica por estrato o ingeniería radica en el hecho de tener la información numérica para posteriores estudios en donde se pueda realizar una “*comparación o correlación*” entre el nivel de habilidad numérica registrado por los estudiantes y el rendimiento en asignaturas como “*calculo y algebra*”, en donde ha habido venido presentando problemas por bajo rendimiento. Se podría, en un estudio posterior para ampliar la validez de las conclusiones del estudio, en donde se mida la “*habilidad de graficación*” y “*habilidad geométricas*” de los estudiantes de ingeniería.

“El análisis estadístico es una herramienta para la toma de decisiones basada en el análisis de la información parcial recolectada”

3.3. Forma de seleccionar una muestra de datos

Es de relevante importancia saber que cuando el “*ingeniero investigador*” se plantea un estudio, este comúnmente se realiza “*de lo particular a lo general*” porque generalmente sólo logra disponer de información parcial del fenómeno investigado o sólo logra disponer de una “*colección de observaciones*” obtenidas de una parte extraída de la población en estudio, es decir, no tiene el conocimiento de todos los resultados derivados de un fenómeno investigado, “*desconoce la población*” o no tiene acceso a la totalidad. En consecuencia, la estadística se va a enfocar en obtener la información a partir de la “*observación o experimentación*” tomando solo una parte de los elementos que forman parte de una población, al observar o crear un experimento del fenómeno poblacional estudiado, con el objetivo “*analizar la información parcial*” contenida en la muestra extraída de la población, con la finalidad de “*generalizar los resultados*” obtenidos a toda la población involucrada en el estudio. Por tanto cuando se plantea el estudio de un fenómeno, su objetivo es la “*obtención de observaciones o resultados*” llámense “*datos*”, a través de la observación parcial del fenómeno o de la creación de un experimento a una “*muestra física de elementos*” obtenidos de la población

física investigada. El listado total de “ n ” observaciones o resultados obtenidos de la muestra física, se le denomina “*muestra de datos*”. El segundo objetivo es el “*análisis exhaustivo*” de la muestra de datos obtenida, a partir de un proceso de “*organización en tablas*”, “*representación gráfica*”, “*cálculo de medidas descriptivas o estimadores*” y la “*aplicación de adecuados métodos de análisis estadístico*”. El tercer objetivo de la estadística es “*emitir conclusiones válidas*” a partir del análisis de la muestra de datos y “*generalizar las conclusiones*” obtenidas, a la población de donde se extrajo la muestra estudiada. Los errores más comunes que se pueden cometer a la hora de tomar una muestra aleatoria son: i) *error de muestreo*, esto ocurre cuando la muestra no es representativa de toda la población. Esto ocurre cuando los valores numéricos de los datos de la población tienen una gran dispersión entre sí, deduciéndose como consecuencia que la muestra tendrá un alto error con respecto a la población de donde se tomó. ii) *error de inferencia*, esto ocurre cuando se hacen conclusiones hacia una población más extensa de la que se tomó originalmente la muestra. A continuación se ilustran las formas de tomar una muestra de datos.

3.3.1. Muestreo aleatorio simple

Es la metodología de muestreo en la que todos los individuos tienen la misma probabilidad de ser seleccionados. La selección de la muestra puede realizarse a través de cualquier mecanismo probabilístico en el que todos los elementos tengan las mismas opciones de salir. Para realizar este tipo de muestreo, y en determinadas situaciones, usualmente se utiliza una tabla de números aleatorios, o también con un ordenador o calculadora generar números aleatorios, comprendidos entre cero y uno, y multiplicarlos por el tamaño de la población, este es el que vamos a utilizar. Dentro de este mecanismo de muestreo se puede considerar: i) *muestreo aleatorio simple sin reposición de los elementos*, ocurre donde cada elemento extraído se descarta para una nueva extracción. Ilustración: si se toma una muestra de bombillos de una población de bombillos producidas por una fábrica para estimar la vida promedio de duración de los bombi-

llos que la integran, no será posible medir la duración más que una vez a cada uno de los bombillos seleccionados en la muestra, dado que no se puede repetir este procedimiento más de una vez con un mismo bombillo. ii) *muestreo aleatorio simple con reposición de los elementos*, en donde las observaciones se realizan con remplazo de los elementos extraídos. Es decir, una vez tomado el elemento, este podría ser tomado nuevamente. Esto puede ocurrir a la hora de hacer encuestas a poblaciones muy grandes, en donde un individuo podría ser encuestado más de una vez. En poblaciones muy grandes, la probabilidad de repetir una extracción es tan pequeña que el muestreo puede considerarse con reposición aunque, realmente, no lo sea.

Ejemplo 7: Considérese una sección de 50 alumnos de ingeniería de la *Universidad Simón Bolívar*. Tome una muestra aleatoria de tamaño 11 y obtenga los datos de las variables: *edad, sexo, # de hermanos, peso, estatura y deporte*.

En este ejemplo se va a considerar como: “*Población Física*” al curso de estadística de la *Universidad Simón Bolívar* ($N = 50$); “*Poblaciones Estadísticas*” a la *Edad, Sexo, # de Hermanos, Peso, Estatura y Deporte*; “*Muestra Aleatoria*” a la parte de la población física seleccionada a través de plan de muestreo o al azar ($n = 11$); y los “*Datos u Observaciones*” a la información obtenida de la muestra sobre algunas características medibles de la población física. La muestra aleatoria se obtuvo a través del mecanismo de la bolsa con los papelitos, supóngase que se obtuvo la muestra de tamaño 11 a partir de la cual se han obtenidos los datos ilustrados en la Tabla 17.

3.3.2. Muestreo aleatorio sistemático

En un muestreo aleatorio sistemático se elige un elemento al azar y a partir de él, a intervalos constantes, se eligen los demás hasta completar la muestra. Es decir, supóngase que se tiene una población de tamaño “ N ” y se quiere tomar una “*muestra sistemática*” de tamaño “ n ”. Entonces se divide a la población de tamaño “ N ” en “ n particiones” organizadas en “ n filas” en donde cada fila tiene “ m elementos”; esto es, una matriz de “ n filas x m columnas”. Luego, se

Tabla 17. Características medibles de la población física

Muestra Aleatoria Física de Tamaño "n = 11"		Características Medibles de la Muestra Física					
Estudiantes de la USB (selección al azar)	X ₁ : Edad	X ₂ : Sexo	X ₃ :# de Hermanos	X ₄ : Peso	X ₅ : Estatura	X ₆ : Deporte	
07 (María): a ₁	17,8	Femenino	3	60,5	1,65	Voleibol	X ₁
19 (Amelia): a ₂	19,3	Femenino	4	64,8	1,70	Natación	X ₂
23 (Luís): a ₃	20,2	Masculino	1	73,3	1,73	Básquet	X ₃
11 (Angélica): a ₄	18,9	Femenino	4	58,9	1,66	Voleibol	X ₄
16 (José): a ₅	20,4	Masculino	2	68,4	1,70	Futbol	X ₅
34 (Jorge): a ₆	18,9	Masculino	3	83,5	1,81	Futbol	X ₆
03 (Alejandra): a ₇	19,4	Femenino	0	68,6	1,73	Maratón	X ₇
45 (Andreina): a ₈	21,9	Femenino	2	61,4	1,55	Básquet	X ₈
33 (Roberto): a ₉	23,5	Masculino	1	75,7	1,67	Natación	X ₉
49 (Marta): a ₁₀	18,9	Femenino	5	50,3	1,63	Voleibol	X ₁₀
17 (Paulina): a ₁₁	20,4	Femenino	2	65,1	1,67	Básquet	X ₁₁

Fuente: Autores.

sortea una observación de la primera fila de la matriz. Supóngase que en el sorteo salió la observación “k” de la primera fila (puesto o ubicación “k” de la primera fila). Entonces la forma de la “muestra sistemática” es: “ $x_1, x_2, x_3, \dots, x_n$ ”; donde $x_1 = O_{1k}, x_2 = O_{2k}, x_3 = O_{3k}, \dots, x_n = O_{nk}$. En este tipo de muestreo hay que organizar la población en “n filas” (particiones de igual tamaño de N) por “m columnas” (# de observaciones de cada partición). Como ilustración se puede exponer que si se tiene una población formada por 100 elementos y queremos extraer una muestra de 25 elementos, en primer lugar debemos establecer el intervalo de selección que será igual a $100/25 = 4$. A continuación elegimos el elemento de arranque, tomando aleatoriamente un número entre el 1 y el 4, y a partir de él obtenemos los restantes elementos de la muestra “n = 2, 6, 10, 14, ..., 98”.



Figura 14. Muestreo aleatorio sistemático (1). Fuente: Autores.

Nótese que al tomar una muestra sistemática de una población “N”, se tiene que: $N = n \times m$ (Tabla 18).

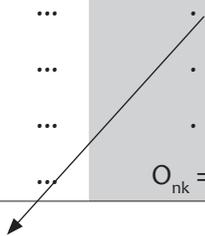
Ejemplo 8: Utilice el muestreo sistemático para tomar una muestra de tamaño 5 ($n = 5$) de una población de 20 ($N = 20$) estudiantes de estadística.

Se divide la población en 5 particiones ($n = 5$) de tamaño 4 ($m = 4$), dado que “ $5 \times 4 = 20$ ”. Luego, de la primera partición o fila

Tabla 18. Muestra sistémica 1

Matriz	Observación en la posición 1	Observación en la posición 2	Observación en la posición 3	...	Observación en la posición k	...	Observación en la posición m
1° Partición	O_{11}	O_{12}	O_{13}	...	$O_{1k} = x_1$...	O_{1m}
2° Partición	O_{21}	O_{22}	O_{23}	...	$O_{2k} = x_2$...	O_{2m}
3° Partición	O_{31}	O_{32}	O_{33}	...	$O_{3k} = x_3$...	O_{3m}
.
.
.
n° Partición	O_{n1}	O_{n2}	O_{n3}	...	$O_{nk} = x_n$...	O_{nm}

Fuente: Autores.



Muestra sistémica

se sortea uno. Supóngase que salió la 2° ($k = 2$) observación de la primera fila. Entonces los elementos de la muestra sistemática son: $x_1, x_2, x_3, \dots, x_n$; donde $x_1 = O_{1k}, x_2 = O_{2k}, x_3 = O_{3k}, \dots, x_n = O_{nk}$.

Tabla 19. Muestra sistémica 2

	Observación en la posición 1	Observación en la posición $k = 2$	Observación en la posición 3	Observación en la posición 4
1° Partición	O_{11}	$O_{12} = x_1$	O_{13}	O_{14}
2° Partición	O_{21}	$O_{22} = x_2$	O_{23}	O_{24}
3° Partición	O_{31}	$O_{32} = x_3$	O_{33}	O_{34}
4° Partición	O_{41}	$O_{42} = x_4$	O_{43}	O_{44}
5° Partición	O_{51}	$O_{52} = x_5$	O_{53}	O_{54}

Fuente: Autores.

Muestra sistemática

“El muestreo sistemático “no es” recomendable si hay cierta periodicidad en las observaciones de la característica que se está analizando; ya que se podría cometer el error de tomar elementos muy similares entre sí, lo que naturalmente no representaría a la población en estudio”.

La Figura 15 ilustra a los elementos de la muestra sistemática en donde hay cierta periodicidad (todos son muy parecidos entre sí, por tanto no representan fielmente a la población de donde se han extraído).

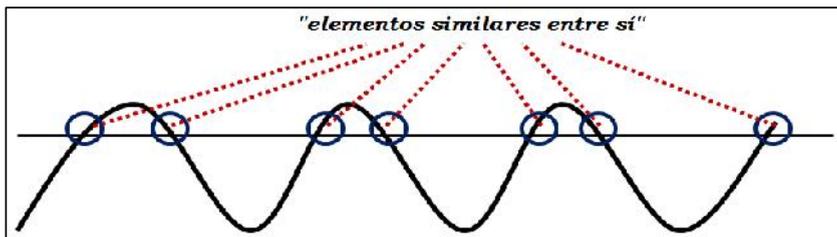


Figura 15. Muestreo aleatorio sistemático (2). Fuente: Autores.

Ejemplo 9: Como ilustración de lo anterior, supóngase que se quiere determinar el promedio de compra diaria en un supermercado. Para ello se calcula el promedio de compra el día 15 de cada mes y se calcula ese promedio todos los 15 de cada mes (*día de pago*). El problema que se presenta con la escogencia de ese día en forma sistemática y consecutiva para calcular el promedio de compra diaria, es que ese día, los trabajadores comunes cobran sus quincenas o sueldos y por lo tanto, el gasto promedio es superior al de los días restantes del mes. De ahí que calcular el promedio de compra tomando como referencia y en forma sistemática el día 15 de cada mes, ha de conllevar a una conclusión errónea del promedio diario de compra en mencionado supermercado.

3.3.3. Muestreo aleatorio estratificado

En un muestreo aleatorio estratificado se requiere de un conocimiento previo de la población con el fin de dividir la misma en particiones o estratos (esto es particionar la población de acuerdo a una característica, por ejemplo: *sexo, nacionalidad, preferencias deportivas, nivel de instrucción, estrato social, gasto mensual en pollo, etc.*). El conocimiento previo de la población permite hacer una clasificación o partición de la población de acuerdo a la característica a ser tomadas en cuenta para conformación de los estratos (*particiones de acuerdo a un criterio*), por ejemplo una partición de acuerdo a una característica llamada “*sexo*”, estaría conformada por “*2 estratos o particiones*”, una para el “*sexo femenino*” y otra para el “*sexo masculino*”. Las divisiones, particiones o estratificaciones de la población se deben realizar de acuerdo a las siguientes reglas: Los elementos que se dejan en un mismo estrato de acuerdo con determinada característica deben ser *homogéneos (parecidos)* y los elementos que se dejan en distintos estratos de acuerdo con determinada característica deben ser *heterogéneos (diferentes)*. Luego, la muestra final estará conformada por todos los elementos elegidos aleatoriamente, en los distintos estratos. Por otro lado, la distribución del tamaño de la muestra de acuerdo al número de elementos pertenecientes a los diferentes estratos se denomina “*afijación*”, y esta puede tomar

diferentes denominaciones: La “afijación simple” ocurre cuando se toma igual número de elementos por estrato, es decir, el número de elementos por estrato es de un mismo tamaño; esto es, si se toma una muestra de 20 elementos, y se tienen “4 estratos”, la muestra tendrá “5 elementos” en cada uno de los “4 estratos”. La “afijación proporcional” ocurre cuando la distribución del número de elementos por estrato se hace tomando un número proporcional de elementos contenidos en cada estrato, es decir, se toman los elementos de la muestra proporcionales al tamaño o número de elementos del estrato. De ahí que, si un estrato tiene el 50% de los elementos de la población, entonces el 50% de los elementos la muestra deben ser de ese estrato. De igual forma si un estrato tiene el 20% de los elementos de la población, entonces el 20% de los elementos la muestra deben ser de ese estrato y la “afijación óptima” ocurre cuando se tiene en cuenta la previsible dispersión de los resultados, de modo que se considera tanto la proporción de elementos por estrato y como también la varianza existente en cada estrato. De ahí que se toma una muestra proporcional al tamaño y a la varianza de cada estrato. La varianza (se refiere a una medida de la dispersión de los elementos del estrato poblacional), un estrato poblacional con varianza grande (con mucha dispersión), requiere el tomar un número de elementos mayor que el de un estrato poblacional con varianza pequeña (con poca dispersión).

Ejemplo 10: Como se debería tomar la muestra si se quieren determinar las preferencias por un determinado candidato en una elección.

Se diseña una ficha técnica que incluye el universo a encuestar (N), la cobertura geográfica donde se tomará la muestra, el tamaño de la muestra (n) y el periodo de tiempo en el que se realizará la encuesta. Ilustrando lo anterior, supóngase que se considera:

- Estrato social: alto-medio-bajo.
- Edad: joven-intermedio-viejo.
- Sexo: femenino-masculino.



Figura 16. Muestreo aleatorio estratificado. Fuente: Autores.

“Ficha Técnica”							
Encuesta realizada a una muestra estratificada de tamaño “n” tomada a una población votante del “Norte de Santander” de tamaño “N” en lapso comprendido entre la primera y tercera semana del mes de agosto del presente año. Se tomarán en cuenta los estratos: <i>Estrato Socio- Económico, Edad, y Sexo.</i>							
Fecha: _____		Entrevista Realizada en: _____					
Estratificación Socio-Económica			Estratificación por Edad			Estratificación por Sexo	
<u>Bajo</u>	<u>Medio</u>	<u>Alto</u>	18 - 30	31 - 45	Más de 45	F	M
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<u>Entrevista</u>							
						SI	No
1.) ¿Cree Ud. En un cambio en la dirigencia regional?						<input type="radio"/>	<input type="radio"/>
2.) ¿Esta Ud. Dispuesto a votar por una nueva dirigencia regional?						<input type="radio"/>	<input type="radio"/>
3.) ¿A quien considera Ud. Como el candidato con mayor opción para dirigir los destinos de la región?							
Candidato A	Candidato B	Candidato C	Candidato D	Candidato E	Otro		
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		

Figura 17. Ficha técnica de un muestreo estratificado. Fuente: Autores.

“Cada una de las personas seleccionadas para ser encuestadas e integrar los elementos de la muestra, está ya predeterminada por el encuestador, es decir, la ficha técnica se llena de antemano antes de la entrevista”.

3.3.4 Muestreo aleatorio por conglomerado (clúster)

En un muestreo aleatorio por conglomerado (*clúster*) opera de los métodos presentados hasta ahora están pensados para seleccionar directamente los elementos de la población, es decir, las unidades muestrales son diversos elementos de la población. En el muestreo por conglomerados se utilizan grupos de elementos de la población a la que llamamos conglomerados. Los estudiantes de las diversas secciones de estadística, una caja de determinado producto, los costales de papa traídos de una finca, las diversas cajas o urnas electorales en un centro de votación, etc., son asumidos como grupos o conglomerados. Cuando los conglomerados son áreas geográficas suele hablarse de “*muestreo por áreas*”. El muestreo por conglomerados consiste en seleccionar aleatoriamente un cierto número de conglomerados (*el necesario para alcanzar el tamaño muestral establecido*) y en investigar después todos los elementos pertenecientes a los conglomerados elegidos.

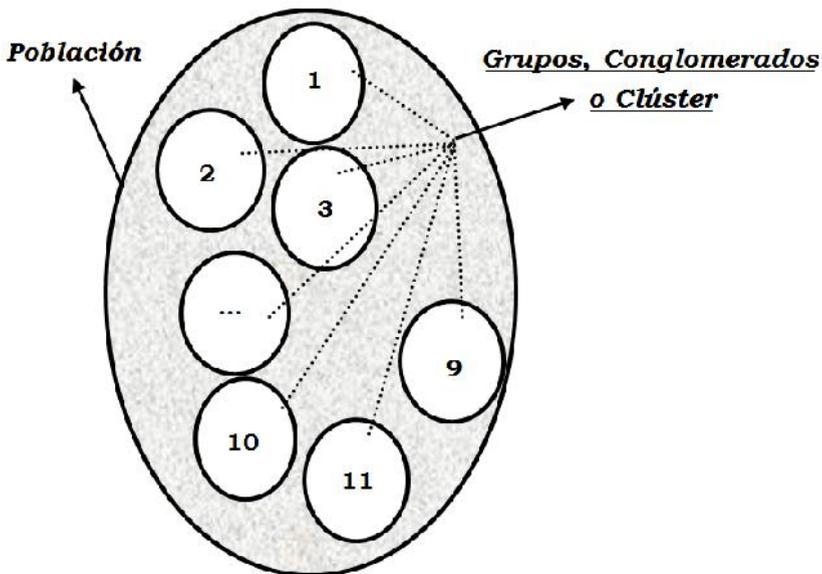


Figura 18. Muestreo aleatorio por conglomerado (clúster). Fuente: Autores.

Supóngase que se tiene una población dividida en “9 conglomerados” similares. Supóngase además que al sortear entre los 9 conglomerados o clúster resultan escogidos el grupo 3 y el grupo 9. Entonces se examinarán los elementos de los grupos 3 y 9. Como los grupos 3 y 9 son parecidos a la población; éstos también se parecen entre sí. Por lo tanto bastaría con analizar completamente, de ser posible, sólo uno de ellos.

En éste muestreo se toman comunidades básicas de trabajo llamados “grupos” o “conglomerados” en donde cada uno de ellos por su similitud representa a la población total. Es decir, cada grupo o conglomerado es muy similar a la población de donde se extrajo pero en pequeño. Supóngase que se tiene una población de “N” elementos la cual se divide en “ $k_1, k_2, k_3, \dots, k_n$ ” conglomerados, si entre estos “ $k_1, k_2, k_3, \dots, k_n$ ” conglomerados se realiza un sorteo y se obtienen algunos grupos. Supóngase “ k_1, k_3 y k_7 ” conglomerados, entonces se estudian los elementos de dichos conglomerados. En este caso se habla de muestreo por conglomerado de una etapa.

Si en la muestra inicial que incluía el estudio de a los elementos de los conglomerados “ k_1, k_3 y k_7 ” se divide en otros subgrupos de conglomerados; esto es, para “ k_1 ” se tiene “ $k_{11}, k_{12}, k_{13}, \dots, k_{1n}$ ”, para “ k_2 ” se tiene “ $k_{21}, k_{22}, k_{23}, \dots, k_{2n}$ ” y para “ k_3 ” se tiene “ $k_{31}, k_{32}, k_{33}, \dots, k_{3n}$ ”; y se realiza un segundo sorteo obteniéndose nuevos subgrupos, sean éstos, “ k_{11}, k_{15} y k_{19} ”; “ k_{21}, k_{25} y k_{29} ”; “ k_{31}, k_{35} y k_{39} ” y se estudian los elementos de dichos conglomerados, se estaría hablando de muestreo de dos etapas. Así sucesivamente, si “ k_{11}, k_{15} y k_{19} ”; “ k_{21}, k_{25} y k_{29} ”; “ k_{31}, k_{35} y k_{39} ” se divide en otros subgrupos de conglomerados y se realiza un tercer sorteo obteniéndose y se estudian los elementos de dichos conglomerados, se habla de muestreo de tres etapas.

Ejemplo 11: Describa cómo tomar una muestra de un cargamento de 100 cajas de manzanas que llegan en un camión para conocer el color, el sabor y la textura de la manzana, de manera se pueda determinar si se acepta o rechaza el cargamento.

Esta población es homogénea ya que los elementos de cada caja se parecen a los de la totalidad de la población. Se puede decir

que cada caja es un “cluster” (pequeño grupo que representa a la población). De ahí que se procede a sortear entre las 100 cajas unas 5 cajas, luego se puede sortear entre cada una de las 5 cajas 2 bandejas de manzanas. Después se procede a sortear entre cada una de las bandeja seleccionadas unas 3 manzanas. Luego, la muestra a examinar es de “ $5 \times 2 \times 3 = 30$ manzanas. Se puede concluir que se ha realizado un muestreo de tres etapas.

“En ciencias sociales se utiliza con frecuencia el muestreo por conglomerado, porque reduce costos en forma considerable”

3.3.5. Muestreo no probabilístico

A veces, para estudios exploratorios, el “muestreo probabilístico” resulta excesivamente costoso y se acude a “métodos no probabilísticos”, aun siendo conscientes de que no sirven para realizar generalizaciones (estimaciones inferenciales sobre la población), pues no se tiene certeza de que la muestra extraída sea representativa ya que no todos los sujetos de la población tienen la misma probabilidad de ser elegidos. En general se seleccionan a los sujetos siguiendo determinados criterios procurando, en la medida de lo posible, que la muestra sea representativa. En algunas circunstancias un muestreo no probabilístico, por ejemplo los estudios de caso-control, donde los casos no son seleccionados aleatoriamente de la población, se usa en los casos epidemiológicos. Entre los métodos de muestreo no probabilísticos más utilizados en investigación encontramos:

Muestreo por cuotas, también denominado en ocasiones “accidental”. Se asienta generalmente sobre la base de un buen conocimiento de los estratos de la población y/o de los individuos más “representativos” o “adecuados” para los fines de la investigación. Mantiene, por tanto, semejanzas con el muestreo aleatorio estratificado, pero no tiene el carácter de aleatoriedad de aquél. En este tipo de muestreo se fijan unas “cuotas” que consisten en un número de individuos que reúnen unas determinadas condiciones, como ilustración supóngase que se tiene 20 individuos de 25 a 40 años, de sexo femenino y residentes en Cúcuta. Una vez determinada la cuota se

eligen los primeros que se encuentren que cumplan esas características. Este método se utiliza mucho en las encuestas de opinión.

Muestreo intencional o de conveniencia, este tipo de muestreo se caracteriza por un esfuerzo deliberado de obtener muestras “representativas” mediante la inclusión en la muestra de grupos supuestamente típicos. Es muy frecuente su utilización en sondeos preelectorales de zonas que en anteriores votaciones han marcado tendencias de voto. También puede ser que el investigador seleccione directa e intencionadamente los individuos de la población. El caso más frecuente de este procedimiento el utilizar como muestra los individuos a los que se tiene fácil acceso (*los profesores de universidad emplean con mucha frecuencia a sus propios alumnos*). *Bola de Nieve*, Se localiza a algunos individuos, los cuales conducen a otros, y estos a otros, y así hasta conseguir una muestra suficiente. Este tipo se emplea muy frecuentemente cuando se hacen estudios con poblaciones “marginales”, delincuentes, sectas, determinados tipos de enfermos, etc. y *muestreo discrecional*, a criterio del investigador los elementos son elegidos sobre lo que él cree que pueden aportar al estudio.

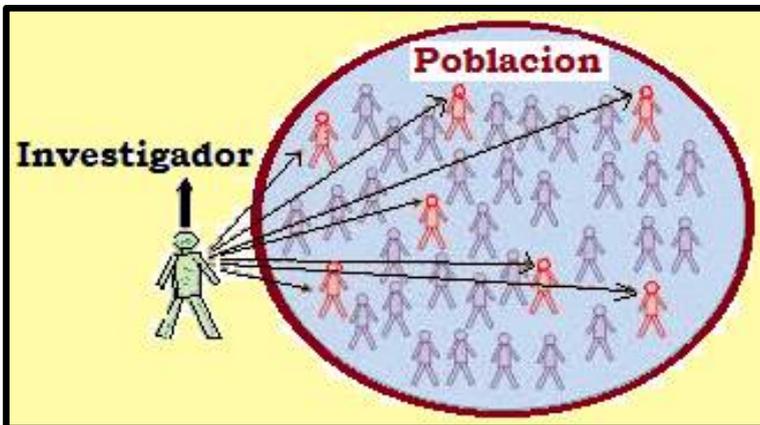


Figura 19. Muestreo intencional o de conveniencia. Fuente: Autores.

“El muestreo más indicado es el que aporte mayor información a menor costo; tomar muestras aleatorias a través de algún mecanismo al azar trae como consecuencia que los errores que se cometen en la escogencia de los elementos de la misma, se compensan; por otro lado,

para trabajar con una muestra se realiza un muestreo y cuando se trabaja con una población se realiza un censo”

3.4. Forma de calcular el tamaño una muestra de datos

Para establecer una ley física, biológica o social respecto a un determinado fenómeno, el investigador debe comenzar con una colección de datos u observaciones. La materia prima de la estadística consiste en conjuntos de datos obtenidos al contar o medir características cuantificables del fenómeno. La recopilación de la información implica tener un especial cuidado de garantizar que la información sea completa y correcta. Uno de los principales problemas que enfrenta el investigador en la recolección de los datos, es cuanta información se ha de requerir para obtener conclusiones lo más valederas posibles, esto se refiere al “*tamaño de la muestra*”. El mismo, está relacionado directamente con la precisión de los resultados que se obtendrán. En términos generales, se puede decir, que cuanto más grande sea el tamaño de la muestra mayor precisión tendrán los resultados que se obtendrán. Un análisis más exhaustivo del tamaño de la muestra nos dice que éste está íntimamente ligado a la variabilidad o dispersión (*Varianza σ^2*) de las observaciones de la población estadística en estudio. Si una población estadística tiene poca varianza (*es aproximadamente homogénea*), basta obtener una *muestra pequeña* para generar conclusiones acertadas *aproximadas* respecto a la población estadística en estudio. Por otro lado, si una población estadística tiene una varianza muy grande (*es heterogénea*), se requerirá obtener una *muestra grande* para generar conclusiones acertadas *aproximadas* respecto a la población estadística en estudio. Lo anterior se puede visualizar mejor a través de las dos ilustraciones siguientes: i) Para obtener conclusiones respecto al nivel de colesterol, triglicéridos, nivel de glóbulos rojos y plaquetas del torrente sanguíneo de un individuo, basta con una pequeña muestra de laboratorio, dado que la sangre es homogénea (*similar en todo el torrente*); ii) Para determinarlas preferencias por un determinado artista de cine, se requerirá de una muestra más grande y estratifica-

da. Ya que la preferencia por el artista dependerá de una cantidad de aspectos tales como: actitud del artista, atractivo físico del artista, identificación del artista con un estilo, habilidades actorales, etc. Los aspectos mencionados están íntimamente relacionados con distintas formas de apreciación. En este caso, los resultados dependerán de la manera como se estratifique la población y el tamaño de los estratos a tomar en cuenta. Como consecuencia, se puede decir que en la ilustración citada, la población es altamente heterogénea. De ahí que los artistas que triunfan sean tan distintos en estilos, físicos y carismas entre sí.

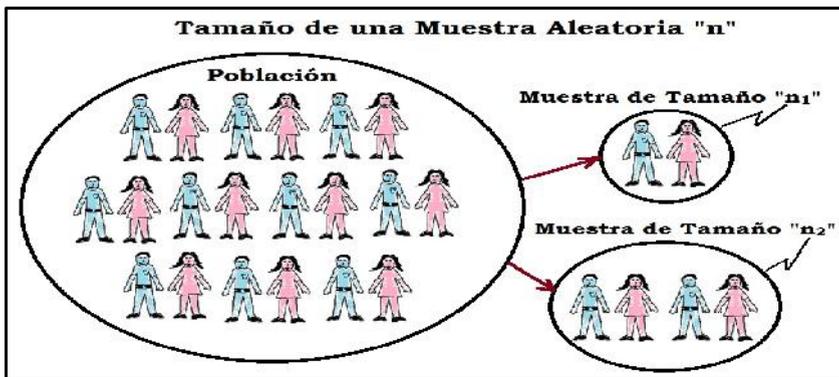


Figura 20. Tamaño una muestra de datos. Fuente: Autores.

Los objetivos de la determinación del tamaño adecuado de una o varias muestra, es por una lado, en el caso de una muestra estimar a partir de la misma, algunos indicadores (parámetros o medidas) que “describan o ilustren la característica medible”, por otro lado, si son varias muestras tomadas con la finalidad de realizar comparaciones entre una o más características de interés, “detectar una determinada diferencia, si realmente existe, entre dos poblaciones en estudio con un cierto grado de garantía”; así por ejemplo, en un estudio de investigación epidemiológica la determinación de un tamaño adecuado de las muestras tendría como objetivo su factibilidad. Si el número de sujetos es insuficiente habría que modificar los criterios de selección, solicitar la colaboración de otros centros o ampliar el período

de reclutamiento, dado que los estudios con tamaños muestrales insuficientes, no son capaces de detectar diferencias entre grupos, llegando a la conclusión errónea de que no existe tal diferencia. Por otro lado, si el número de sujetos es excesivo, el estudio se encarece desde el punto de vista económico y humano. Por otro lado es la estimación de parámetros poblacionales con un determinado margen de error prefijado por el investigador. Algunas fórmulas generales para calcular el tamaño de una muestra aleatoria son:

1) *Tamaño de la muestra de una población finita cuantitativa de tamaño conocido:*

Fórmula para calcular el tamaño de una “muestra aleatoria n” en donde se tiene el tamaño de la población “N”:

$$n = \frac{N \times \sigma^2 \times Z_t^2}{(N - 1) \times \delta^2 + \sigma^2 \times Z_t^2}; \text{ donde:}$$

n: es el tamaño de la muestra a ser seleccionada a través de un mecanismo aleatorio.

N: es el tamaño de la población (número total de elementos involucrados en el estudio).

σ : es la desviación estándar de la población, cuando no se tiene su valor, generalmente suele utilizarse un valor constante de 0,5 o se calcula un estimador “S” a partir de los datos de una muestra piloto.

Z_t : es un valor de la curva normal estándar N (0,1), que depende del nivel de confianza que asigne. El nivel de confianza indica la probabilidad de que los resultados de nuestra investigación sean ciertos. Un 95 % de confianza es lo mismo que decir que nos podemos equivocar con una probabilidad del 5%. Los valores de Z_t más utilizados de acuerdo a ciertos niveles de confianza son:

Tabla 20. De nivel de confianza

Valor de Z_t	1,15	1,28	1,44	1,65	1,96	2,24	2,58
Nivel de Confianza	75%	80%	85%	90%	95%	97,5%	99%

Fuente: Autores.

(Por tanto, si se pretende obtener un nivel de confianza del 95% necesitamos poner en la fórmula $k = 1,96$).

δ : representa el error muestral aceptado en un estudio y que se asigna arbitrariamente, este se define como diferencia existente entre los resultados producto de analizar los datos de una muestra aleatoria y los resultados producto de analizar todos los datos de una población. Así por ejemplo: i) si los resultados de una encuesta dicen que 100 personas comprarían un producto y aceptamos error muestral del $\pm 5\%$ estamos diciendo que las personas que compraran el producto estará entre 95 y 105 personas (100 con un error $\delta = \pm 5$); ii) si se hace una encuesta de satisfacción laboral a los empleados de una empresa con un error muestral de $\pm 3\%$, en donde el resultado final muestra que el 60% de los empleados encuestados se muestran satisfechos, significa que el porcentaje de empleados satisfechos de la empresa se encuentra entre el 57% y el 63% (60% con un error $\delta = \pm 3\%$); iii) si los resultados de una encuesta electoral indicaran que un candidato iba a obtener el 55% de los votos y el error muestral asignado fue del $\pm 4\%$, se prevé que el porcentaje real de votos estará entre el 51% y el 59% (55% con un error $\delta = \pm 4\%$). Suele utilizarse un valor δ que varía entre el 1% ($\delta = 0,01$) y 9% ($\delta = 0,09$). En definitiva, el valor lo asignará o lo pondrá arbitrariamente el investigador.

Ejemplo 12: Calcule el tamaño de una “muestra aleatoria n ” a tomar de una población estadística numérica de 70.000 elementos con un nivel de confianza del 95%, en donde se sabe que la desviación estándar es 0,5 y se escoge arbitrariamente a $\delta = \pm 0,05$.

Se tiene que: $N = 70000$ y dado que el nivel de confianza es de 95%, se tiene que $Z_t = 1,96$. Además, $\delta = \pm 0,05$ y $\sigma = 0,5$. Entonces el tamaño de la muestra de acuerdo a la ecuación queda:

$$n = \frac{N \times \sigma^2 \times Z_t^2}{(N-1) \times \delta^2 + \sigma^2 \times Z_t^2} = \frac{(70000) \times (0,5)^2 \times (1,96)^2}{(70000-1) \times (0,05)^2 + (0,5)^2 \times (1,96)^2} = 382,07 \approx 382$$

II) *Tamaño de la muestra de una población finita cuantitativa de tamaño desconocido:* Fórmula para calcular el tamaño de una “muestra aleatoria n ” en donde no se tiene el tamaño de la población “ N ”:

$$n = \frac{\sigma^2 \times Z_t^2}{\delta^2}; \text{Donde:}$$

n : es el tamaño de la muestra a ser seleccionada a través de un mecanismo aleatorio.

σ : es el mismo valor definido en la fórmula anterior.

Z_t : es el mismo valor definido en la fórmula anterior.

δ : es el mismo valor definido en la fórmula anterior.

Ejemplo 13: En un proyecto de investigación en una zona rural, se ha calculado que cerca del 30% de los niños padecen de malnutrición crónica. Este dato se basa en estadísticas nacionales sobre malnutrición en las zonas campesinas. Calcule el tamaño de una “muestra aleatoria n ” a tomar de una población de tamaño desconocido con un nivel de confianza del 95%, en donde el margen de error aceptado es $\pm 0,05$.

Se tiene que: $p = 0,3$ y que por tanto $q = 1 - p = 1 - 0,3 = 0,7$. Dado que el nivel de confianza es de 95%, se tiene que $Z_t = 1,96$. Además, $\delta = \pm 0,05$. Entonces el tamaño de la muestra de acuerdo a la ecuación queda:

$$n = \frac{Z_t^2 \times p \times q}{\delta^2} = \frac{(1,96)^2 \times (0,3) \times (0,7)}{(0,05)^2} = 322,69 \approx 323$$

Ejemplo 14: Supóngase que un investigador desea describir las estaturas en pulgadas de los estudiantes de un curso de 50 estudiantes de “Ingeniería de Sistemas” de la “Universidad Simón Bolívar”. Construya una Tabla de “Distribución de Frecuencias”

Se procede a realizar las mediciones de las estaturas de cada uno de los 50 ($n = 50$) estudiantes del curso en pulgadas. Supóngase que las mediciones obtenidas fueron las siguientes:

Tabla 21. Estaturas de los estudiantes de la U.S.B

	61	63	65	63	69	67	53	58	60	65
Estaturas de los	64	65	64	72	68	66	55	57	66	62
“50” estudiantes	62	65	64	71	68	66	56	59	61	64
de Estadística	63	65	63	62	67	60	57	59	61	62
de la USB.	64	64	63	69	67	66	58	60	61	70

Fuente: Autores.

Antes de construir la tabla de distribución de frecuencias se plantean dos preguntas: ¿Cómo obtener el número de clases? y ¿de qué longitud debe ser cada clase?

Para obtener el “# de clases (k)” se usa la fórmula de Sturges:
 $K = 1 + 3,322 \times \text{Log}_{10}(n)$

K: # de clases.

n: # de observaciones o datos.

$\text{Log}_{10}(n)$: Logaritmo en base 10 del número de datos considerados para el estudio.

Para obtener la “longitud de cada clase” se usa la fórmula:

$$l_{ic} = \frac{A}{K}$$

l_{ic} : Longitud del intervalo de clase.

A: Amplitud, donde $A = V_M - V_m$.

V_M : Valor mayor o más grande.

V_m : Valor menor o más pequeño.

K: # de clases.

“En otro caso, usted mismo puede asignar el número de clases, el cual se recomienda este comprendido entre 5 y 15 clases, cuando se tenga un número de datos mayor o igual a 30”.

Continuando con el **Ejemplo 17**, para obtener el intervalo de clase usando la fórmula de Sturges, se tiene: $K = 1 + 3,322 \times \text{Log}_{10}(50) = 6,64 \approx 7$.

Por otro lado, la longitud del intervalo de clase queda: $l_{ic} = \frac{19}{7} = 2,7 \approx 3 \Rightarrow l_{ic} \approx 3$. Nótese que $A = V_M - V_m = 72 - 53 = 19$.

Para construir los intervalos de la tabla de “Distribución de Frecuencias”, se procede de la siguiente manera: Se inicia colocando el “menor valor ($V_m = 53$)” en el “límite inferior (L_i)” de la “Clase 1”:

Luego, a este valor puesto llamado “límite inferior (L_i)” de la “Clase 1” ($V_m = 53$)” se le suma la “longitud del intervalo de clase ($l_{ic} \approx 3$)” ($53 + 3 = 56$) y se obtiene el “límite inferior (L_i)” de la “Clase 2”.

De manera análoga, al “límite inferior (L_i)” de la “Clase 2” se le suma la “longitud del intervalo de clase ($l_{ic} \approx 3$)” ($56 + 3 = 59$) y se obtiene el “límite inferior (L_i)” de la “Clase 3”.

Y así sucesivamente hasta completar todos los “límites inferiores (L_i)” de las 7 clases:

Una vez que se han construido todos los intervalos en la columna de “rec

Recuento, se procede a tabular o a marcar una rayita (/) o conteo, por cada uno de los 50 datos u observaciones de la variable “estatura” en el intervalo y/o clase que lo contenga o que le corresponda. Realizado lo anterior se rellenan por clase, los cuadros de las columnas referentes a: frecuencia (f_i), frecuencia acumulada (F_i), marca de clase (X_i), frecuencia relativa ($\frac{f_i}{n}$), la frecuencia relativa acumulada ($\frac{F_i}{n}$), porcentaje de la frecuencia relativa ($\frac{F_i}{n} \times 100$), y porcentaje de la frecuencia relativa acumulada.

La Tabla 22 muestra la distribución de frecuencia del Ejemplo 14.

Tabla 22. Distribución de frecuencia de las estaturas de los estudiantes de la U.S.B.

Clases	Intervalos $L_i - L_s$	Recuento	f_i	F_i	X_i	$\frac{f_i}{n}$	$\frac{F_i}{n}$	$\left(\frac{f_i}{n}\right) \times 100$	$\left(\frac{F_i}{n}\right) \times 100$
Clase 1	53 – 55	//	2	2	54	0.04	0.04	4	4
Clase 2	56 – 58	////	5	7	57	0.10	0.14	10	14
Clase 3	59 – 61	////////	9	16	60	0.18	0.32	18	32
Clase 4	62 – 64	//////////	15	31	63	0.30	0.62	30	62
Clase 5	65 – 67	//////////	12	43	66	0.24	0.86	24	86
Clase 6	68 – 70	////	5	48	69	0.10	0.96	10	96
Clase 7	71 – 73	//	2	50	72	0.04	1.00	4	100

Fuente: Autores.

A modo de observación se puede notar que:

$f_3 = 9$ es la frecuencia de la “Clase 3”.

$F_4 = 31$ es la frecuencia acumulada de la “Clase 4”.

$X_5 = 66$ es la marca de clase de la “Clase 5”.

$n = F_k = F_7 = \sum_{i=1}^k f_i = \sum_{i=1}^7 f_i = 50$ es la frecuencia acumulada de la “Clase 4”.

I) *Representación gráfica mediante Histogramas*: Es un gráfico de barras continua (éstas barras son rectángulos), donde la base de cada rectángulo es el intervalo de clase y la altura es la frecuencia de la clase. Para que exista continuidad en las barras del histograma, se utilizarán los “límites reales (L.R.)” de cada clase. En el caso de “datos que corresponden a números enteros”, esta continuidad se genera sumando y restando a los “límites de cada clase ($L_i - L_s$)” el valor 0,5. Es decir: Los “límites reales” se generan mediante de intervalos de la forma: $(LR_i - LR_s) = (L_i + 0,5 - L_s - 0,5)$.

“Si se trabaja con números con un decimal (Ejemplo: 31,7 – 45,8 – 12,9 – 37,4 – ...) Los “límites reales” se generarán a través de intervalos de la forma: $(LR_i - LR_s) = (L_i + 0,05 - L_s - 0,05)$ ”

“Si se trabaja con números con dos decimales (Ejemplo: 31,74 – 45,85 – 12,98 – 37,43 – ...) Los “límites reales” se generarán a través de intervalos de la forma: $(LR_i - LR_s) = (L_i + 0,005 - L_s - 0,005)$ ”

II) *Representación gráfica mediante polígono de frecuencias*: Es un “diagrama de línea” que se obtiene uniendo los puntos medios superiores de cada rectángulo del histograma mediante semi-rectas. Cada punto tiene como coordenada horizontal a la “marca de clase (X_i)” y como coordenada vertical, la “frecuencia (f_i)” de la clase. Se acostumbra iniciarlo en la marca de clase inmediatamente anterior a la primera y concluirlo en la marca de clase inmediatamente posterior a la última.

Ejemplo 18: Realice el histograma y el polígono de frecuencias correspondiente a la *Tabla de distribución de frecuencia del “Ejemplo 17”*.

Para construir el *histograma* y el *polígono de frecuencia* es necesario construir una tabla que contenga las *clases*, los *intervalos* de clase, los *límites reales* de cada intervalo, la *frecuencia* de cada clase y la *marca de clase*.

Tabla 23. Límites reales del histograma y el polígono de frecuencias

Clases	Intervalos $L_i - L_s$	Límites Reales $LR_i - LR_s$	f_i	X_i
Clase 1	53 – 55	52.5 – 55.5	2	54
Clase 2	56 – 58	55.5 – 58.5	5	57
Clase 3	59 – 61	58.5 – 61.5	9	60
Clase 4	62 – 64	61.5 – 64.5	15	63
Clase 5	65 – 67	64.5 – 67.5	12	66
Clase 6	68 – 70	67.5 – 70.5	5	69
Clase 7	71 – 73	70.5 – 73.5	2	72

Fuente: Autores.

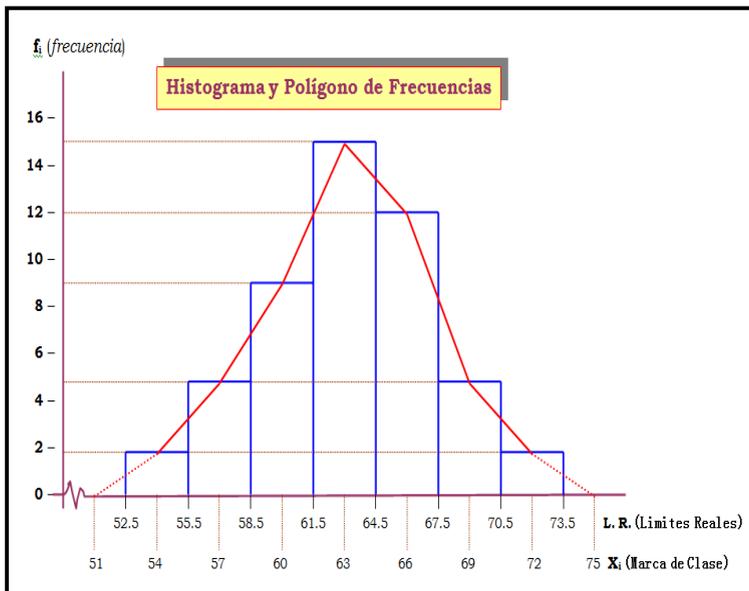


Figura 21. Histograma y el polígono de frecuencias. Fuente: Autores.

3.5. Estimadores poblacionales

El objetivo más importante de una investigación es generalizar los resultados obtenidos de una muestra aleatoria a la población de donde se extrajo la muestra, es decir, hacer una generalización de los resultados muestrales. Como las poblaciones se describen mediante medidas numéricas denominadas *parámetros poblacionales* (es una medida numérica resumen, que se calcula usando todos los datos de la población, tales como el promedio poblacional " μ ", la varianza poblacional " σ^2 ", la proporción poblacional " p ", ...), es de vital importancia estimar o acercarse al parámetro poblacional a través de un *estimador poblacional* (medida numérica resumen que se calcula usando los datos de la muestra aleatoria, tales como el promedio muestral " \bar{X} ", la varianza muestral " S^2 ", la proporción muestral " \hat{P} ", ...). De lo anterior, se puede deducir que un estimador poblacional es una medida numérica que se calcula a una muestra aleatoria con la finalidad de aproximarse de alguna manera a los parámetros poblacionales. Existen muchos tipos de estimadores, están los de "tendencia central", están los de "variabilidad o dispersión" están los de "posición o ubicación" y los de "porcentaje". los estimadores más importantes son: la media (tendencia central), la mediana (posición), la varianza (variabilidad) y la proporción de éxitos (porcentaje).

3.5.1. Medidas de tendencia central

Al describir colecciones de observaciones, con frecuencia es conveniente resumir la información con un solo número, si este valor tiende a situarse hacia el centro de la distribución o colección de datos, éste se denomina "*medida de tendencia central*" o de centralización, dado que aportan información respecto al "*centro físico*" o "*centro de gravedad*" de una colección de datos de una variable en estudio. A estos valores se les llama parámetros cuando se calculan a partir de la población total y estadísticas o estimadores cuando se calculan a partir de los datos de una muestra. Las medidas de tendencia central son "*la media, la mediana y la moda*" y sirven como puntos de referencia para interpretar los valores de un conjunto de datos.

l) *La media aritmética* (\bar{X}): Es aquella medida que representa un promedio o centro físico de los datos, aunque la misma no está necesariamente en la mitad. Ésta se puede calcular cuando la variable en estudio es numérica. La media es uno de los valores más representativos de una distribución de datos, cuando la distribución es aproximadamente simétrica. Como primera ilustración, supóngase que un estudiante obtiene 30 puntos en una prueba de matemática. Este puntaje, por sí mismo tiene muy poco significado, pero si calificación promedio de los estudiantes que presentaron la prueba fue de 18 puntos, se puede decir que la calificación del alumno se ubica notablemente sobre el promedio. Pero si la calificación promedio de los estudiantes que presentaron la prueba fue de 45 puntos, entonces la conclusión sería muy diferente, debido a que se ubicaría muy por debajo del promedio del curso. En otras palabras, para que un valor tenga significado hay que contar con valores de referencia generalmente relacionados con este valor. Como segunda ilustración, supóngase que se quiere emprender un negocio de venta de empanadas. Un aspecto importante a conocer es la determinación del grado de aceptación de las empanadas que se han de vender. Un mecanismo es conocer si, en promedio, a los consumidores les gustará la sazón o sabor de las empanadas. Para ello se toma una muestra aleatoria de 25 personas y se les da a probar la empanada pidiendo que le asignen una calificación del 1 al 10 a la calidad de la sazón o sabor de la empanada. A partir de la calificación aportada por los consumidores, la cual fue cuantificada a través de los números del 1 al 10, se puede calcular un promedio para tener una medida que indique el grado de aceptación del producto. La determinación de un promedio marcará la tendencia de aceptación o rechazo del producto, lo cual será determinante para conocer el futuro éxito del negocio.

Para calcular la “media aritmética” (\bar{X}), se usa la expresión:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

x_i : representa el “i-ésimo” dato de la colección.

n : representa el número de datos.

“La media se denota a través de \bar{X} cuando se calcula a partir de los datos de una muestra y se llama en forma general, estadística o estimador poblacional; pero cuando se calcula a partir de la población total, se denota a través de μ y se le llama en forma general, parámetro poblacional”

II) La mediana (\tilde{X}): Es el valor de posición central de un conjunto de observaciones que están ordenadas en forma creciente o decreciente. Es decir, una colección de datos con la forma:

" $x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)}$ " (Sucesión de observaciones o datos organizadas en orden creciente de magnitud). La mediana divide el conjunto de observaciones en dos partes iguales con el 50% de los datos cada una. De ahí que el 50% de los datos son menores que la mediana y el otro 50% son mayores que la mediana.

Para calcular la “mediana” (\tilde{X}), en una colección de datos se comienza organizando las observaciones de menor " $x_{(1)}$ " a mayor " $x_{(n)}$ " (o de forma creciente) " $x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)}$ ". Luego: i) Si el número de observaciones es impar, la mediana es el valor central de este ordenamiento; pero ii) Si el número de observaciones es par, la mediana es el promedio de los dos valores centrales de este ordenamiento.

“La mediana es usada como indicador central cuando la media no resulta ser un buen indicativo del centro físico de los datos, por haber sido afectada por algunas observaciones o datos extremadamente altos o bajos fuera del contexto numérico regular que presenta la distribución”

III) La moda (\hat{X}): es una medida que está muy relacionada con su nombre “moda” (la tendencia que más se repite), numéricamente se define como la observación que más se repite o que se da con mayor frecuencia. Aunque, si todas las variables tienen la misma frecuencia se dirá que no hay moda.

Para calcular la “moda” (\hat{x}), en una colección de datos se comienza organizando las observaciones de menor " $x_{(1)}$ " a mayor " $x_{(n)}$ " (o de forma creciente) " $x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)}$ ". Luego, a partir de este ordenamiento es posible ubicar visualmente el valor numérico que más se repite.

“Ejemplo 19”: Considérese la variable “edad” en años de la muestra de 11 estudiantes de estadística ($n = 11$), de la tabla de datos del **Ejemplo 7**.

Edad	17,8	19,3	20,2	18,9	20,4	18,9	19,4	21,9	23,5	18,9	20,4
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}

Calcule la media (\bar{x}), la mediana (\tilde{x}) y la moda (\hat{x}).

“Cálculo de la media (\bar{x})”

Edad ($n = 11$)	17,8	19,3	20,2	18,9	20,4	18,9	19,4	21,9	23,5	18,9	20,4
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}

$$\bar{x} = \frac{\sum_{i=1}^{11} x_i}{11} = \frac{17,8 + 19,3 + 20,2 + 18,9 + 20,4 + 18,9 + 19,4 + 21,9 + 23,5 + 18,9 + 20,4}{11}$$

$$\Rightarrow \bar{x} = \frac{219,6}{11} = 19,96364 \Rightarrow \bar{x} \approx 20.$$

Se concluye que la “edad promedio” de la muestra de los 11 estudiantes es de aproximadamente 20 años.

“Cálculo de la mediana (\tilde{x})”

Edad											
(Orden	17,8	18,9	18,9	18,9	19,3	19,4	20,2	20,4	20,4	21,9	23,5
Creciente)											
	$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$	$x_{(11)}$

Como $n = 11$ es impar, $\hat{X} = 18,9$ la mediana es el dato central
 $x_{(6)} = 19,4$

Se concluye que el valor central de la muestra de las “edades” de los 11 estudiantes es 19,4.

“Cálculo de la moda (\hat{x})”

Edad											
(Orden	17,8	18,9	18,9	18,9	19,3	19,4	20,2	20,4	20,4	21,9	23,5
Creciente)											
	$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$	$x_{(11)}$

Nótese que al ordenar los datos en forma creciente se puede observar que la moda es $\hat{X} = 18,9$ dado que es el dato que más se repite.

3.5.2. Medidas de variabilidad o dispersión

Una medida de variabilidad es un indicador del grado de dispersión (*acercamiento o alejamiento*) de un conjunto de observaciones de una variable en torno a la media o centro físico de la misma. Si la dispersión es poca, entonces la medida de variabilidad es pequeña y esto significa que *los datos están muy apiñados en torno a su centro físico*, si la medida de variabilidad es grande, entonces hay mucha dispersión entre los datos y esto significa que *los datos están muy alejados o dispersos en torno su centro físico*. Por otro lado, si todos los datos son iguales, entonces la medida de variabilidad es cero. Entre más grande sea el grado de “dispersión” de los datos una población (*esto es menor uniformidad o mayor heterogeneidad de las*

observaciones), entonces será menor la representatividad o confiabilidad tendrán los estimadores obtenidos de las muestras extraídas de estas poblaciones. Por el contrario, si el grado de “dispersión” de los datos una población es pequeño (esto es *mayor uniformidad o mayor homogeneidad de las observaciones*), entonces será mayor la representatividad o confiabilidad tendrán los estimadores obtenidos de las muestras extraídas de estas poblaciones. Cuando el grado de “dispersión” es cero quiere decir que todos los datos son iguales. Nótese que el grado de dispersión va depender de la unidad de medida utilizada. Por otro lado, la importancia de una medida de variabilidad radica en que si se tiene una población muy poco dispersa u homogénea, entonces se requerirá tomar muestras más pequeñas para explicar el comportamiento de la misma, pero si se tiene una población muy dispersa o heterogénea, entonces el tamaño de la muestra a tomar para explicar el comportamiento de la misma deberá ser más grande y aplicando una buena estrategia de muestreo.

1) *Amplitud o rango (A)*: Es una medida que describe la longitud del recorrido de la variable. En este recorrido están todos los valores que puede tomar la variable en estudio. La amplitud describe la distancia existente entre la menor observación y la mayor observación. Esta medida se define como la diferencia entre el mayor valor y el menor valor de una colección de datos y se usa para describir el alcance de las variaciones extremas tales como: *tiempo máximo – tiempo mínimo, precio máximo de acciones – precio mínimo de acciones, calificación mayor – calificación menor, ... ,etc.* Para calcular la amplitud, se usa la ecuación:

$$A = V_M - V_m \quad V_M : \text{Mayor valor de la distribución de datos.}$$

$$V_m : \text{Menor valor de la distribución de datos.}$$

II) *Varianza* (S^2) y *desviación estándar* (S): La *varianza* es otra alternativa a la *desviación media absoluta* que también nos proporciona un indicativo de la forma como se distribuyen los datos en torno a la media. Ésta se calcula a través del promedio de los cuadrados de las desviaciones con respecto a la media (*también se puede calcular con respecto a la mediana*). La *desviación estándar* es la raíz cuadrada de la *varianza* $S = +\sqrt{S^2}$.

Para calcular la “*varianza*” (S^2), se usa la expresión:

$$S^2 = \frac{1}{n-1} \times \left(\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n} \right)$$

x_i : representa el “*i-ésimo*” dato de la colección.

La *desviación estándar* (S) es la raíz cuadrada de la *varianza* $S = +\sqrt{S^2}$ y es interpretada cuando los datos tienen la forma de una curva de datos aproximadamente simétrica, de la siguiente manera:

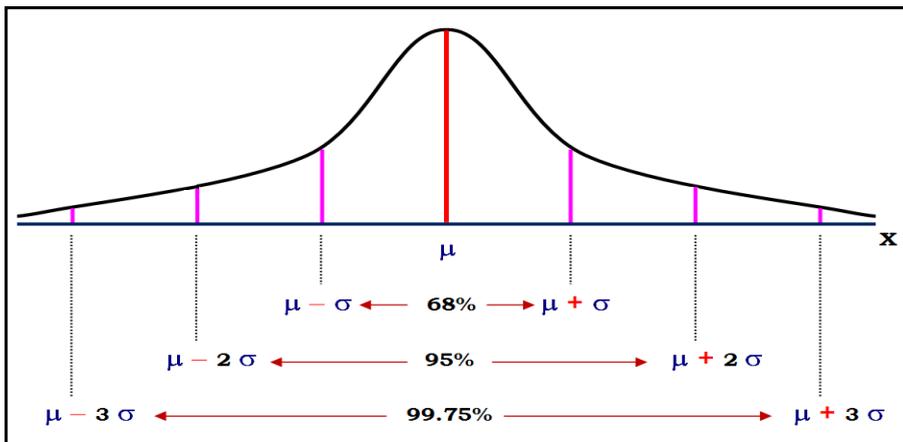


Figura 22. Ilustración de la desviación estándar en la curva normal.
Fuente: Autores.

Ejemplo 20: Considérese la variable “edad” en años de la muestra de 11 estudiantes de ingeniería de sistemas ($n = 11$), de la Tabla de datos del “Ejemplo 7”.

Edad	17,8	19,3	20,2	18,9	20,4	18,9	19,4	21,9	23,5	18,9	20,4
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}

Calcule la varianza (s^2) y la desviación estándar (s).

“Cálculo de la varianza (s^2)”

Edad ($n = 11$)	17,8	19,3	20,2	18,9	20,4	18,9	19,4	21,9	23,5	18,9	20,4
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}

Sabiéndose que “ $n = 11$ ”. La varianza se calcula por la ecuación:

$$S^2 = \frac{1}{11-1} \left(\sum_{i=1}^{11} x_i^2 - \frac{\left(\sum_{i=1}^{11} x_i \right)^2}{11} \right)$$

en la que al calcularse por partes se tiene:

$$\sum_{i=1}^{11} x_i^2 = (17,8)^2 + (19,3)^2 + (20,2)^2 + (18,9)^2 + (20,4)^2 + (18,9)^2 + (19,4)^2 + (21,9)^2 +$$

$$(23,5)^2 + (18,9)^2 + (20,4)^2 = 4409,54$$

$$\sum_{i=1}^{11} x_i = 17,8 + 19,3 + 20,2 + 18,9 + 20,4 + 18,9 + 19,4 + 21,9 + 23,5 + 18,9 + 20,4 = 219,6$$

Luego:

$$S^2 = \frac{1}{11-1} \left(\sum_{i=1}^{11} X_i^2 - \frac{\left(\sum_{i=1}^{11} X_i \right)^2}{11} \right) = \frac{1}{11-1} \left(4409,54 - \frac{(219,6)^2}{11} \right) \Rightarrow S^2 = 2,55.$$

Se concluye que la varianza de la muestra de las “edades” de los 11 estudiantes es 2,55.

“Cálculo de la desviación estándar (S)”

La varianza se calcula por la ecuación:

$$S = +\sqrt{S^2} = +\sqrt{2,55} = 1,5968 \Rightarrow S = 1,6$$

Se concluye que la desviación estándar de la muestra de las “edades” de los 11 estudiantes es 1,6.

3.5.3. Medidas de posición (cuartiles)

En una distribución de los datos de una variable, es importante saber determinar la *posición relativa de los datos en términos porcentuales*. Los “cuantiles” representan un conjunto de “medidas de posición” que dividen a una distribución en términos de porcentaje y permiten encontrar los datos ubicados en cualquier posición planteada en estos términos. Para esclarecer lo anterior, considérese lo siguiente: *un investigador diseña un examen para estudiantes que aspiran entrar en la facultad de arquitectura. El examen mide las habilidades en el diseño y distribución de espacios. Para ello, el investigador toma una muestra aleatoria de 45 aspirantes a la carrera y les aplica el examen. Una vez que el investigador tiene los resultados del examen, los ordena ya sea en forma creciente o en algún tipo de tabla de frecuencias. Una vez organizados los datos, el investigador podría plantearse interrogantes tales como: ¿qué puntaje obtenido en el examen representa el puntaje central, es decir, el puntaje que divide las observaciones en dos partes iguales 50% por arriba y 50% por debajo?, ó ¿qué puntaje obtenido en el examen corresponde al tope del 25% de las calificaciones más deficiente en la prueba, y a su vez, sobre qué puntaje obtenido en el examen se establece el piso o se da inicio de las calificaciones que están*

por encima del 80% de los mejores puntajes obtenidos en la prueba? Para responder a estas interrogantes, se calculan algunos Cuantiles o medidas de posición. Los Cuantiles más usados, son los cuartiles.

Se entiende por cuartil " Q_i " al valor numérico que divide una distribución ordenada en forma creciente en "cuatro partes iguales" en términos de porcentaje. Los cuartiles son tres (3) y se ilustran a continuación.

El 25% de las observaciones es \leq a Q_1

El 50% de las observaciones es \leq a Q_2

El 75% de las observaciones es \leq a Q_3

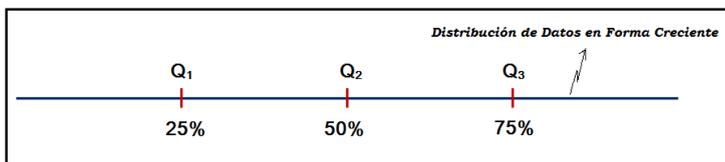


Figura 23. Cuartiles. Fuente: Autores.

Fórmulas para calcular los "cuartiles" Q_i :

Cálculo de los "cuartiles" en una colección de datos de cualquier naturaleza numérica: Se ordenan los datos en forma creciente " $x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)}$ " y se ubica la posición del cuartil a través de la ecuación adjunta.

$$PQ_i = \frac{i \times (n + 1)}{4} = x_{(i)}; \text{ donde } i = 1, 2, 3.$$

PQ_i : representa la posición del "i-ésimo" cuartil

n : representa el número de datos.

$x_{(i)}$: representa al valor del "i-esimo" cuartil.

Ejemplo 21: Considérese la variable "edad" en años de la muestra de 11 estudiantes de estadística ($n = 11$), de la tabla de datos del "Ejemplo 7".

Edad	17,8	19,3	20,2	18,9	20,4	18,9	19,4	21,9	23,5	18,9	20,4
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}

Calcule: Q_1 , Q_2 y Q_3 .

Se ordenan los datos en forma creciente:

Edad											
(Orden	17,8	18,9	18,9	18,9	19,3	19,4	20,2	20,4	20,4	21,9	23,5
Creciente)											
	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$	$X_{(8)}$	$X_{(9)}$	$X_{(10)}$	$X_{(11)}$

“Cálculo del Q_1 ”: Se ubica la posición del “cuartil 1” a través de la ecuación $PQ_1 = \frac{1 \times (11+1)}{4} = 3$. Luego, dado que la posición del “ Q_1 ” es “3” ($PQ_1 = 3$); se tiene el “cuartil 1” es $Q_1 = x_{(3)} = 18,9$ $Q_1 \Rightarrow = 18,9$

“Cálculo del Q_2 ”: Se ubica la posición del “cuartil 2” a través de la ecuación $PQ_2 = \frac{2 \times (11+1)}{4} = 6$. Luego, dado que la posición del “ Q_2 ” es 6 ($PQ_2 = 6$); se tiene que el “cuartil 2” es: $Q_2 = x_{(6)} = 19,4 \Rightarrow Q_2 = 19,4$

“Cálculo del Q_3 ”: Se ubica la posición del “cuartil 3” a través de la ecuación $PQ_3 = \frac{3 \times (11+1)}{4} = 9$. Luego, dado que la posición del “ Q_3 ” es “9” ($PQ_3 = 9$); se tiene que el “cuartil 3” es: $Q_3 = x_{(9)} = 20,4 \Rightarrow Q_3 = 20,4$

“Interpretación”

- El 25% de los estudiantes de la muestra tiene edades \leq a 18,9 años.
- El 50% de los estudiantes de la muestra tiene edades \leq a 19,4 años.
- El 75% de los estudiantes de la muestra tiene edades \leq a 20,4 años.

Nota: Si la posición de un cuartil es un número como: 1,5; 2,5; 3,5; 4,5; ...; 10,5; es decir, está exactamente entre dos posiciones, por ejemplo, la posición "3,5" está entre $x_{(3)}$ y $x_{(4)}$ entonces se calcula un promedio entre ambos valores, para el caso del ejemplo anterior

$$\text{se tiene: } x_{(i)} = \frac{x_{(3)} + x_{(4)}}{2} = \frac{18,9 + 18,9}{2} = 18,9$$

El cuartil correspondiente a la posición "3,5" es "18,9".

3.5.4. Gráfico boxplot o de caja

Este gráfico consiste en cercar las observaciones a través de los cuartiles Q_1 y Q_3 que delimitan la caja, las cuales son medidas de posición que encierran al 50% de las observaciones. La caja está partida en dos partes (no necesariamente iguales) a través de una línea que representa a la mediana \tilde{X} . Luego, se han de considerar como observaciones extrañas o atípicas, a aquellas observaciones o datos que sobrepasen el cerco de delimitado inferiormente por cualquiera de las tres expresiones: " $Q_1 - 1,0 \times (Q_3 - Q_1)$ " ó " $Q_1 - 1,5 \times (Q_3 - Q_1)$ " ó " $Q_1 - 2,0 \times (Q_3 - Q_1)$ " y superiormente por cualquiera de las tres las expresiones: " $Q_3 + 1,0 \times (Q_3 - Q_1)$ " ó " $Q_3 + 1,5 \times (Q_3 - Q_1)$ " ó " $Q_3 + 2,0 \times (Q_3 - Q_1)$ ".

Lo anterior es ilustrado en la Figura 24.

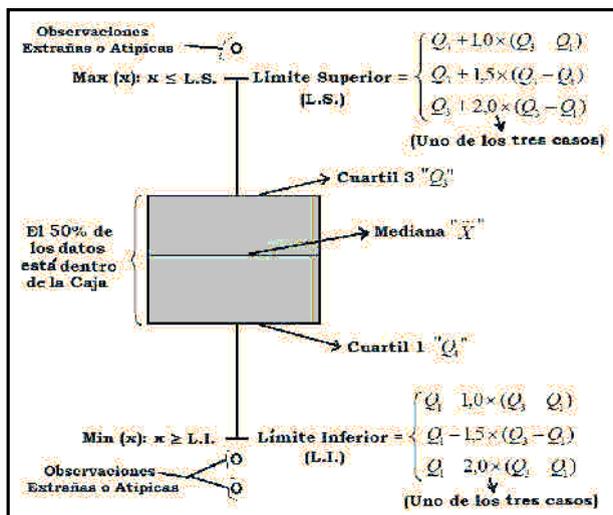


Figura 24. Gráfico boxplot o de caja 1. Fuente: Autores.

Ejemplo 22: Utilice los datos no tabulados del “Ejemplo 17” que refiere a las estaturas en pulgadas de los estudiantes de un curso de 50 estudiantes de “Estadística” de la Universidad Simón Bolívar, para realizar un gráfico de “Caja o Boxplot”.

Desarrollo

Tabla 24. Estatura de 50 estudiantes

Estaturas de los “50” estudiantes de estadística de la USB	61	63	65	63	69	67	53	58	60	65
	64	65	64	72	68	66	55	57	66	62
	62	65	64	71	68	66	56	59	61	64
	63	65	63	62	67	60	57	59	61	62
	64	64	63	69	67	66	58	60	61	70

Fuente: Autores.

Anteriormente, a estos datos se les calculó: $Q_1 = 60$ pulgadas, $\bar{x} = 63$ pulgadas y $Q_3 = 66$ pulgadas.

El límite inferior (L.I.) queda con alguna de las tres expresiones siguientes:

$$"Q_1 - 1,0 \times (Q_3 - Q_1) = 60 - 1,0 \times (66 - 60) = 54"$$

$$"Q_1 - 1,5 \times (Q_3 - Q_1) = 60 - 1,5 \times (66 - 60) = 51"$$

$$"Q_1 - 2,0 \times (Q_3 - Q_1) = 60 - 2,0 \times (66 - 60) = 48"$$

El límite superior (L.S.) queda con alguna de las tres expresiones siguientes:

$$"Q_1 + 1,0 \times (Q_3 - Q_1) = 60 + 1,0 \times (66 - 60) = 66"$$

$$"Q_1 + 1,5 \times (Q_3 - Q_1) = 60 + 1,5 \times (66 - 60) = 69"$$

$$"Q_1 + 2,0 \times (Q_3 - Q_1) = 60 + 2,0 \times (66 - 60) = 72"$$

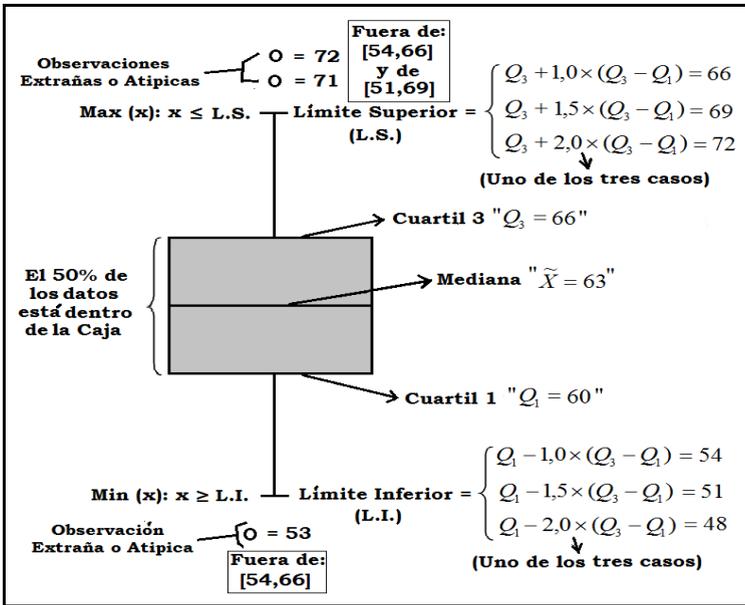


Figura 25. Gráfico boxplot 2 o de caja de las estaturas en pulgadas.

3.5.5. Proporción de éxitos (\hat{p})

Es una medida probabilística que mide en forma porcentual, a partir de una muestra de datos, la posibilidad de ocurrencia de una de las alternativas de un fenómeno. Se calcula a través de la fórmula:

$$\hat{p} = \frac{\# \text{ de Éxitos de Muestra}}{\# \text{ Total de Elementos de la Muestra}}$$

Ejemplo 23: ¿De qué manera, un investigador podría estimar o determinar la proporción de fumadores de los clientes que asisten a una tasca?

Para determinar la proporción de fumadores que asisten a una tasca, el investigador podría tomar una *muestra aleatoria (tomada al azar)* de 9 clientes de la tasca un día cualquiera. Supóngase que los resultados obtenidos fueron los ilustrados en la Tabla 25.

Tabla 25. Fumadores y no fumadores de la tasca

Muestra Física de Tamaño 9 (n = 9): Clientes de la Tasca	X Fumar (Característica) (X(a) = x)
Cliente 1: a ₁	X(a ₁) = x ₁ = Fuma
Cliente 2: a ₂	X(a ₂) = x ₂ = No Fuma
Cliente 3: a ₃	X(a ₃) = x ₃ = No Fuma
Cliente 4: a ₄	X(a ₄) = x ₄ = Fuma
Cliente 5: a ₅	X(a ₅) = x ₅ = Fuma
Cliente 6: a ₆	X(a ₆) = x ₆ = Fuma
Cliente 7: a ₇	X(a ₇) = x ₇ = No Fuma
Cliente 8: a ₈	X(a ₈) = x ₈ = Fuma
Cliente 9: a ₉	X(a ₉) = x ₉ = Fuma

Fuente: Autores.

Nota: La expresión " $X(a) = x$ " se lee "fumador o no fumador"

La proporción de fumadores de la muestra es:

$$\hat{p} = \frac{\# \text{ de fumadores}}{\# \text{ total de encuestados}} = \frac{6}{9} = 0,67$$

Es decir, se puede calcular a partir de la muestra tomada una proporción \hat{p} de fumadores igual a un 0,67 ($\hat{p} = 0,67$) es decir, un 67%. El proceso de inferencia va a consistir en lograr aproximarse a la verdadera proporción de fumadores que asiste regularmente a la tasca a partir de la información obtenida en la muestra de los 9 fumadores.



Figura 26. Porcentaje de fumadores y no fumadores. Fuente: Autores.

Actividades del capítulo III

1. ¿Cuál es la diferencia entre una población física y una población estadística?
2. ¿Cuál es la diferencia entre una muestra física y una muestra estadística?
3. ¿Qué es y con qué propósito se toma una muestra aleatoria física?
4. ¿Qué es una muestra aleatoria de datos?
5. ¿A qué se le llama muestrear?
6. ¿En qué se asemejan y en que discrepan los conceptos estadísticos población y muestra?
7. Defina e ilustre las diferentes escalas de medición utilizadas.
8. ¿En qué se parecen y en qué discrepan la estadística descriptiva y la estadística inferencial?
9. Formule y plante usted, el análisis estadístico de un problema particular.
10. ¿Cómo enfoca la estadística sus problemas?
11. ¿De qué se vale la estadística para organizar y representar adecuadamente las observaciones, datos o resultados obtenidos?
12. Defina lo que significa dato estadístico.
13. Defina lo que significa variables estadística.
14. Considérese la tabla anexa:

Tamaño	Textura	Ubicación	Peso	Clasificación	Tipos
23.5	Lisa	Profunda	60,5	Rocosa	2
30.8	Rugosa	Intermedia	64,8	No Rocosa	3
25.7	Rugosa	Superficial	73,3	Rocosa	3
23.2	Lisa	Superficial	58,9	No Rocosa	5
20.4	Lisa	Intermedia	68,4	No Rocosa	3
28.9	Rugosa	Superficial	83,5	Rocosa	6
31.3	Lisa	Profunda	68,6	Rocosa	3
24.1	Áspera	Profunda	61,4	Rocosa	6
23.6	Rugosa	Intermedia	75,7	No Rocosa	3
26.0	Lisa	Profunda	50,3	Rocosa	1
21.3	Áspera	Superficial	65,1	No Rocosa	2

De acuerdo a esa Tabla, responda las preguntas siguientes:

- Determine cuales variables son cuantitativas y cuales variables son cualitativas.
 - Determine cuales variables son continuas y cuales variables son discretas.
15. Describa la diferencia entre poblaciones físicas y poblaciones estadísticas.
 16. Dé ejemplos de poblaciones físicas y poblaciones estadísticas.
 17. Diga a qué se llama muestrear.
 18. ¿Cuáles son los objetivos de la aplicación de un buen método de muestreo?
 19. Describa paso a paso el mecanismo que usted utilizaría, para tomar una muestra aleatoria de tamaño “n” de la población finita de tamaño “N” de estudiantes de ingeniería de la Universidad Simón Bolívar, para aplicarles un test de habilidades abstractas y razonamiento geométrico en el plano, para determinar un número que represente el promedio de estos estudiantes en estas habilidades.

20. Utilice el muestreo sistemático para tomar una muestra de tamaño 25 de una población de 500 estudiantes de Ingeniería de la Universidad Simón Bolívar.
21. ¿Cuáles son las ventajas y desventajas del muestreo sistemático?
22. ¿Cómo tomaría usted la muestra estratificada de una parte de la población de la Universidad Simón Bolívar, para determinar las preferencias electorales por un determinado grupo de candidatos aspirantes a un cargo importante en dicha institución? Describa paso a paso el procedimiento a utilizar y construya la ficha del test a ser aplicado a la muestra predeterminada por usted. Indique los estratos a ser considerados es la elaboración de la ficha test.
23. ¿De qué manera se deben elegir los estratos en un muestreo estratificado?
24. Indique a su parecer, cuáles son las ventajas y desventajas del muestreo estratificado.
25. Indique la forma de tomar la muestra para determinar un número indicativo del promedio de compra diario en un supermercado importante de la ciudad.
26. Describa cómo tomar una muestra sistemática de un cargamento de 500 cajas de mangos que llegan a un supermercado para conocer el estado de los mismos y poder determinar si se acepta o rechaza el cargamento.
27. ¿En qué ciencias es recomendado el muestreo por conglomerado?
28. ¿Dónde no se recomienda el muestreo por conglomerado?
29. ¿En función de que se escoge el mecanismo de muestro más indicado?
30. ¿Cuándo se realiza un muestreo y cuando se realiza un censo?

31. Una vez que se ha tomado la muestra y se tiene la información a ser cuantificada y tabulada, responda las preguntas planteadas a continuación:
- ¿Cuándo es recomendable emplear la escala nominal? Ilústrelo con un ejemplo.
 - ¿Cuándo es recomendable emplear la escala ordinal? Ilústrelo con un ejemplo.
 - ¿Cuándo es recomendable emplear la escala de Intervalos? Ilústrelo con un ejemplo.
 - ¿Cuándo se es recomendable emplear la escala de razón? Ilústrelo con un ejemplo.
32. Suponga que se les realiza un “test de geometría” a 50 aspirantes a estudiar Ingeniería de Sistemas en la Universidad Simón Bolívar. Los resultados fueron los siguientes:

74	44	49	33	34	33	76	58	68	39
29	41	45	32	83	58	73	47	40	26
37	47	69	53	55	58	49	45	65	47
54	50	51	66	80	73	57	61	56	50
38	45	51	44	41	68	45	93	43	17

Construya una “*tabla de distribución de frecuencias*” que contenga: “clases o Categorías” que posean “límites o valores extremos” ($[L_i - L_s]$), “tabulación o conteo de datos”, “frecuencia” (f_i), “frecuencia acumulada” (F_i), “marca de clase” ($x_i = \frac{L_i + L_s}{2}$), “frecuencia relativa” ($\frac{f_i}{n}$), “frecuencia relativa acumulada” ($\frac{F_i}{n}$), “porcentaje de la frecuencia relativa” ($\frac{f_i}{n}$). $_{100}$ y “porcentaje de la frecuencia relativa acumulada” ($\frac{F_i}{n}$). $_{100}$. Por otro lado:

- Construya un *histograma* y el *polígono de frecuencias*.

- Construya el *histograma de frecuencias acumuladas y la ojiva*.

33. En un examen de arquitectura de la Universidad Simón Bolívar de una sección de 40 estudiantes, se obtuvieron los siguientes resultados:

88	74	67	49	69	38	86	77
97	67	78	69	84	50	39	58
93	77	98	75	98	77	64	78
65	67	75	84	73	56	78	89
72	62	74	81	79	81	86	78

Represente estos datos a través de tres (3) gráficos “*box-plot o de caja*”. Para los siguientes casos:

$$\text{L.I.} = "Q_1 - 1,0 \times (Q_3 - Q_1)" \text{ y L.S.} = "Q_1 + 1,0 \times (Q_3 - Q_1)"$$

$$\text{L.I.} = "Q_1 - 1,5 \times (Q_3 - Q_1)" \text{ y L.S.} = "Q_3 + 1,5 \times (Q_3 - Q_1)"$$

$$\text{L.I.} = "Q_1 - 2,0 \times (Q_3 - Q_1)" \text{ y L.S.} = "Q_3 + 2,0 \times (Q_3 - Q_1)"$$

Sistema informático biomédico inteligente en una TD

El diagnóstico de “*enfermedades hepáticas*” por laboratorios clínicos, es una parte fundamental en el descubrimiento temprano de esta enfermedad para poder ser tratada con eficiencia. Muchas de las herramientas y métodos de diagnóstico existentes de esta enfermedad son muy acertados, pero suelen ser muy tardíos en el momento de realizar el diagnóstico. En los últimos años la investigación y desenvolvimiento de técnicas de “*inteligencia artificial (IA)*”, así como, la rapidez de los procesadores actuales, han abierto nuevos horizontes de aplicación de estas técnicas en problemas que necesitan un cierto grado de inteligencia y autonomía para ser solucionados, un ejemplo claro son las aplicaciones biomédicas a través de “*tarjetas de desarrollo (TD)*”. En este tipo de aplicaciones se busca emitir un diagnóstico médico del estado de salud de un paciente en tiempo real basado en ciertos indicadores presentados por el mismo. El desarrollo de “*sistemas informáticos (SI)*” basado en técnicas de inteligencia artificial implementado sobre una tarjeta de desarrollo para el diagnóstico por laboratorio clínico de pacientes hepáticos puede servir para identificar y clasificar con rapidez y economía las enfermedades hepáticas.

En la actualidad este proceso de detección y diagnóstico en los sistemas de salud en el “*Norte de Santander*” Colombia, puede ser lento, engorroso y en algunos casos traumático, hecho que motiva el “*desarrollo de un sistema informático basado en técnicas de inteligencia artificial implementado sobre una tarjeta de desarrollo para el diag-*

nóstico inmediato de pacientes hepáticos a través exámenes realizados por laboratorio clínico”, que logren una detección rápida y oportuna, brindando una solución a bajos costos a aquellas comunidades con riesgo de este tipo enfermedades. En este trabajo se presenta un sistema computacional inteligente que permite diagnosticar patologías como la “colecistitis hepática” y la “cirrosis hepática” en tiempo real y con un mínimo de error, que diagnóstica enfermedades hepáticas previa realización de exámenes sanguíneos a través de un laboratorio clínico. El trabajo fue desarrollado en dos etapas: 1.) *La primera*, fue la recolección de datos y la selección de la técnica de inteligencia artificial, a ser utilizados en la entrada del sistema y procesamiento adecuado para obtener mayor velocidad de cómputos. 2.) *La segunda*, fue referente a la construcción del sistema inteligente en la que se utilizaron las bibliotecas “*numpy, pybrain, matploit, y sklearn*”; para el proceso de entrenamiento y selección del modelo clasificador. Las metodologías usadas fueron: i) *metodología enfocada a la minería de datos*, y ii) *metodología enfocada en el desarrollo del software*. Ambas metodologías fueron implementadas sobre una tarjeta de desarrollo.

4.1. Tarjetas de desarrollo TD

Una tarjeta de desarrollo es, bajo la perspectiva ingenieril, una herramienta para diseño y prototipado rápido de sistemas digitales o analógicos, que se presenta como un elemento muy útil para el mejoramiento de los procesos de diseño debido a disminución del tiempo de validación de los diseños así como la posibilidad que ofrece de ser una solución y un producto final. La integran un microcontrolador y un entorno de desarrollo (IDE), diseñado para facilitar el uso de la electrónica en proyectos multidisciplinarios. El hardware consiste en una placa de circuito impreso con un microcontrolador, usualmente Atmel AVR, y puertos digitales y analógicos de “*entrada/salida*”, los cuales pueden conectarse a placas de expansión (shields) que expanden las características de funcionamiento de la tarjeta. Por otro lado, el software consiste en un entorno de desarrollo (IDE) basado en el

entorno de processing y lenguaje de programación basado en wiring, así como en el cargador de arranque (bootloader) que es ejecutado en la placa. El microcontrolador de la placa se programa a través de un computador, haciendo uso de comunicación serial mediante un convertidor de niveles RS-232 a TTL serial. Las aplicaciones biomédicas sobre tarjetas de desarrollo y las ventajas que estas puedan ofrecer va a depender de las características y condiciones del problema a ser solucionado dado que cada tarjeta de desarrollo tiene particularidades que ofrecen más o menos beneficio que va a depender de la configuración del problema a ser solucionado. Las tarjetas de desarrollo consideradas para el proyecto fueron tres: la tarjeta de Arduino, la tarjeta de Raspberry Pi y la tarjeta de Beaglebone Black. Después de un análisis de sus alcances y limitaciones, se eligió la tarjeta Beaglebone Black, debido a que esta tarjeta combina el poder de procesamiento de la Raspberry pi y la flexibilidad para conexión de varias interfaces externas de la plataforma Arduino, lo que la hace más completa y está más dirigida a usuarios avanzados, tiene instalado el SO Linux Angstrom, por lo que al igual que la Raspberry pi se pueden usar como un ordenador independiente si se desea. Además se puede instalar en ella gran variedad de SO. A pesar de ser más costosa que la Arduino y Raspberry pi es más beneficiosa en las aplicaciones, esto debido a que tiene incorporadas varias entradas y salidas en la cual se pueden instalar sensores y actuadores externos, por eso es más conveniente para proyectos que interactúan con el mundo real.



Figura 27. Tarjeta de desarrollo. Fuente: productos Raspberry.

4.2. Análisis la base de datos de las enfermedades hepáticas

El análisis de la “*Base de Datos Hepática*” utilizada en la investigación realizada por J.Chacon y R. Miranda en el 2015, para la creación de un sistema informático basado en técnicas de inteligencia artificial implementado sobre una tarjeta de desarrollo para el diagnóstico por laboratorio clínico de pacientes hepáticos, comienza con la descripción de la misma a través de una tabla de datos, la cual consta de 536 patrones de pacientes (filas). Hay cuatro tipos de trastornos hepatobiliares (*Tipos*) y nueve síntomas (*Variables*). Los trastornos hepatobiliares son:

- Tipo 1: *Deterioro (ADL)*
- Tipo 2: *Hepatoma Primario (PH)*
- Tipo 3: *Cirrosis Ligera (CL)*
- Tipo 4: *Cholelithiasis o Colelitiasis (C)*

Los nueve síntomas corresponden a las nueve características de entrada (resultados de diferentes pruebas bioquímicas) las cuales son:

- Variables 1: *Transaminasa glutámico oxalacética (GOT)*
- Variables 2: *Alanina aminotransferasa (GPT)*
- Variables 3: *Deshidrogenasa láctica (LDH)*
- Variables 4: *Gamma glutamil transferasa (GGT)*
- Variables 5: *Nitrógeno ureico en la sangre (BUN)*
- Variables 6: *Volumen corpuscular medio glóbulos rojos (MCV)*
- Variables 7: *Hemoglobina corpuscular media (HCM)*
- Variables 8: *Bilirrubina total (TBIL)*
- Variables 9: *Creatinina (CRTNN)*

En la Tabla 26, se muestra parte de la “*Base de Datos Hepática*” construida con la información de la data de “*Cirrosis Ligera y Colelitiasis*”, debido a que esos dos trastornos son los más indicados de las cuatro enfermedades.

Tabla 26. Base de datos hepática

Resultados de diferentes pruebas bioquímicas y trastorno										
Pacientes	GOT	GPT	LDH	GGT	BUN	MCV	HCM	TBIL	CRTNN	Trastorno
Paciente 1	58	25	360	11	12,9	99,4	33,6	4,8	1,6	Tipo 3
Paciente 2	128	37	414	224	7,1	88,1	29,5	3,4	1,3	Tipo 3
Paciente 3	22	10	282	13	20,7	93,3	31,6	0,4	1,2	Tipo 4
Paciente 4	113	79	374	222	11,8	95,4	31,7	3,4	0,9	Tipo 4
Paciente 5	50	16	654	13	20,5	105,3	35,6	6,9	1,0	Tipo 3
Paciente 6	80	88	234	545	14,6	94,4	31,5	1,3	0,9	Tipo 4
Paciente 7	127	37	414	224	7,1	88,1	29,5	3,4	1,3	Tipo 3

Fuente: Autores.

En la Tabla 27 se muestra el patrón de un paciente específico parte de la data el cual presenta una enfermedad hepática (*Tipo 3: cirrosis ligera*), en donde es mostrada la clase a la que pertenece la enfermedad y las nueve características que describen el patrón hepático.

Tabla 27. Patrón de un paciente específico

Paciente	GOT	GPT	LDH	GGT	BUN	MCV	HCM	TBIL	CRTNN	Trastorno
Paciente 5	50	16	654	13	20,5	105,3	35,6	6,9	1,0	Tipo 3

Fuente: Autores.

En la Tabla 28, cada una de estas características tiene valores de normalidad:

Tabla 28. Valores normales de los síntomas o variables

Características	Unidades	Valores normales	
Transaminasa glutámico oxalacética (GOT)	Unidades/litro	12 - 38 u/l	
Alanina aminotransferasa (GPT)	Unidades/litro	7 - 41 u/l	
Deshidrogenasa láctica (LDH)	Unidades/litro	105 - 333 u/l	
Gamma glutamil transferasa (GGT)	Unidades/litro	9 - 58 u/l	
Nitrógeno ureico en la sangre (BUN)	Miligramos/decilitro	7 - 20 mg/dl	
Volumen corpuscular medio glóbulos rojos (MCV)	Metros cúbicos	79 - 93,3 m ³	
Hemoglobina corpuscular media (HCM)	Picogramos/célula	26,7 - 31,9 pg/célula	
Bilirrubina total (TBIL)	Miligramos/decilitro	0,3 - 1,9 mg/dl	
Creatinina (CRTNN)	Miligramos/decilitro	Hombres	Mujeres
		0,6 - 1,2 mg/dl	0,5 - 0,9 mg/dl

Fuente: Autores.

Se comenzó realizando un pre-procesamiento de la base de datos hepática, lo cual permitió tener una base de datos robustos y sin ninguna clase de ruido que pueda llegar a generar algún tipo de error inesperado. Para ello, fue necesario realizar los siguientes pasos:

- *Limpieza de datos*: eliminar datos ruidosos o inconsistentes (desafío: tratar con datos ruidosos y de baja calidad).
- *Integración de datos*: llevar los datos a un mismo tipo de formato (desafío: tratar con datos no estructurados).
- *Selección de datos*: elegir los datos adecuados (desafío: tratar con grandes volúmenes de datos).

Luego, se hizo un análisis de “*exclusividad / solapamiento*”, para tener una mayor exactitud y comprobar si un patrón de una clase es similar o igual al patrón de otra clase, y de esa manera verificar que no haya una inconsistencia que implique que un mismo dato pertenezca dos clases diferentes. Esto se ilustra en las dos tablas adjuntas donde es mostrado un ejemplo de solapamiento, el cual será considerado para evitar problemas al momento de realizar el diagnóstico. En la Tabla 29 está la información del examen efectuado a un paciente y el resultado numérico del mismo, y en la Tabla 30, exámenes vector línea 2.

Tabla 29. Toma de exámenes (Vector línea 1)

Examen	Resultado
Transaminasa glutámico oxalacética (GOT)	17
Alanina aminotransferasa (GPT)	4
Deshidrogenasa láctica (LDH)	334
Gamma-glutamil transferasa (GGT)	4
Nitrógeno ureico en sangre (BUN)	11,5
Volumen Corpuscular Medio Glóbulos Rojos (MCV)	90,0
Hemoglobina corpuscular media (HCM)	31,5
Bilirrubina Total (TBIL)	0,7
Creatinina (CRTNN)	0,8

Fuente: Autores.

En la Tabla 30 está el patrón de los exámenes del paciente configurado como un vector línea.

Tabla 30. Toma de exámenes (Vector línea 2)

Examen Clínico								
17	4	334	4	11,5	90	31,5	0,7	0,8

Fuente: Autores.

Luego, si el paciente al ser sometido a la evaluación de dos especialistas de la salud, dio como resultado que para el “especialista 1: cirrosis hepática” y para para el “especialista 2: colelitiasis”; a esto se le llama solapamiento.

El estudio de base de datos hepática dio como resultado un porcentaje de solapamiento total de 7,45%; por tanto la base de datos requirió de un análisis de “exclusividad / solapamiento”, y el resultado de este análisis mostro que no hay patrones repetidos en la data hepática; pero como cada fila contiene nueve columnas y estas a su vez contienen uno o dos datos similares a otra fila, existe una probabilidad poco significativa de solapamiento, lo que permite determinar que la data hepática es confiable, dado que no se está clasificando un mismo dato en dos tipos de clases diferentes.

4.3. Técnica de inteligencia artificial utilizada

Para determinar la técnica de inteligencia artificial a utilizar para el diagnóstico de enfermedades hepáticas, se estudiaron los “algoritmos de inteligencia artificial” de las “máquinas de soporte vectorial (SVM)” y de la “red neuronal artificial (RNA)”.

Las máquinas de soporte vectorial (SVM) son estructuras de aprendizaje basadas en la teoría estadística del aprendizaje que se basan en transformar el espacio inicial de entrada, en otro de dimensión superior (infinita) en el que el problema puede ser resuelto mediante un hiperplano óptimo de máximo margen. La formulación de las máquinas de vectores se basa en el principio de minimización

estructural del riesgo que ha demostrado ser superior al principio de minimización del riesgo empírico y presentan un buen rendimiento al generalizar en problemas de clasificación pese a no incorporar conocimiento específico sobre el dominio, este concepto se encuentra expresado en la teoría básica de las “SVM” en problemas de clasificación. La idea principal en esta formulación es construir una función clasificadora que:

- Minimice el error en la separación de los objetos dados (error en clasificación).
- Maximice el margen de separación (mejorar la generalización del clasificador).

Las “SVM” tienen algunas propiedades destacables: 1.) El entrenamiento de una “SVM” es básicamente un problema de programación cuadrática (QP) convexa, que es atractivo por dos motivos: i) Su eficiente computación, esto significa que la mayoría de paquetes de software que solucionan estos problemas son rápidos y precisos, y ii) La garantía que ofrecen de encontrar un extremo global de la superficie de error (*nunca alcanzará mínimos locales*). Esto significa que la solución obtenida es única y óptima para los datos de entrenamiento dados. 2.) A la vez que minimiza el error de clasificación en el entrenamiento, maximiza el margen para mejorar la generalización del clasificador. 3.) No tiene el problema de sobre entrenamiento (*overfitting*) como ocurre con las “redes neuronales”. 4.) La solución no depende de la estructura del planteamiento del problema. 5) Permite trabajar con relaciones no lineales entre los datos (*genera funciones no lineales, mediante kernel*). El producto escalar de los vectores transformados se puede sustituir por el “kernel” por lo que no es necesario trabajar en el espacio extendido. 6.) Generaliza muy bien con pocas muestras de entrenamiento. 7.) Los datos son divididos “*inteligentemente*” en dos partes de diferentes formas y una “SVM” es entrenada para cada forma de división.

Pese a su eficiencia, las “SVM” tienen algunas limitaciones: 1.) La elección de un núcleo adecuado es todavía un área abierta de investigación (*no existe una metodología clara para su elección*). Una vez

elegido el núcleo, los clasificadores basados en “SVM” tienen como único parámetro a ajustar por el usuario es la penalización del error C. 2.) La complejidad temporal y espacial, tanto en el entrenamiento como en la evaluación, son también una limitación. Es un problema sin resolver el entrenamiento con grandes conjuntos de datos (*del orden de millones de vectores soporte*). Los algoritmos existentes para resolver esta familia de problemas tardan un tiempo que depende cuadráticamente del número de puntos. 3.) La “SVM” siempre soluciona un problema bloque, esto significa que un cambio en los patrones de entrenamiento supone obtener una nueva “SVM” pues pueden surgir distintos vectores soporte (aunque existen ya algunas alternativas para el entrenamiento on-line de “SVM” capaces de encajar modificaciones del conjunto de datos sin necesidad de re-entrenar el sistema).

Una vez expuestos las propiedades y limitaciones del uso de las “SVM”, es de gran importancia saber el comportamiento de los datos de la “*base de datos hepática*” cargados antes del entrenamiento, que servirá como punto de partida para la elección del “kernel” (*es el software que constituye la parte más importante del sistema operativo*) y la función a utilizar para crear un modelo de clasificación más eficiente, dado que en el momento de realizar el entrenamiento a través de una “SVM” es necesario elegir la función “kernel”, para mapear los datos de entrenamiento en el espacio del núcleo. A continuación se ilustran la representación de cada uno de los “kernel” que se pueden emplear para clasificar usando la caja de herramientas (Toolbox MSV LS) de Matlab 12 libre y a su vez se muestran cada una de las gráficas generadas durante el entrenamiento.

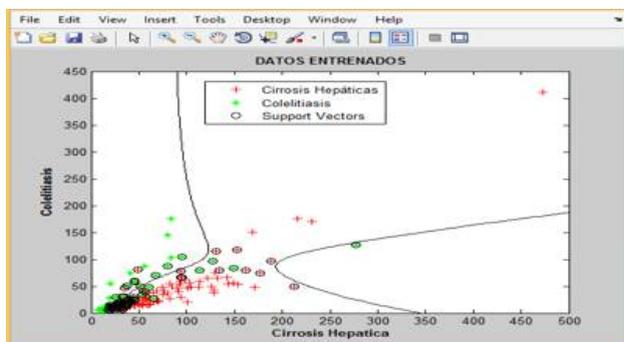


Figura 28. Datos cargados y clasificados usando el Kernel cuadrático.
Fuente: Autores.

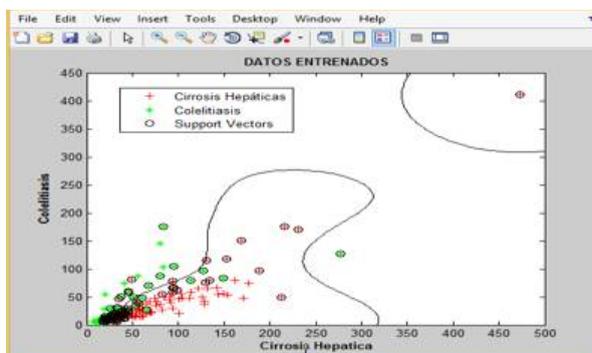


Figura 29. Datos cargados y clasificados usando el Kernel polinómico.
Fuente: Autores.

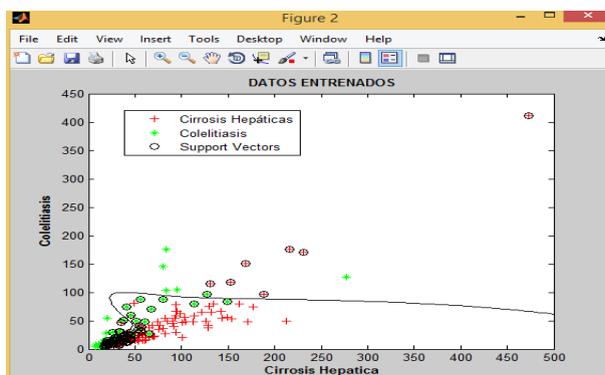


Figura 30. Datos Cargados y clasificados usando el Kernel perceptrón (MLP).
Fuente: Autores.

Al momento de elegir el tipo de kernel para que sea el más eficiente en la utilización del entrenamiento de “SVM” es necesario utilizar un tipo de prueba en la cual se hallara el error con respecto al hiperplano de separación. Los métodos utilizados son:

- “QP” - Programación cuadrática.
- “SMO” - Mínimo de optimización secuencial.
- “LS” - Mínimos cuadrados.

La desviación de dos clases es una medida de error de clasificación binaria que tiene la forma mostrada en la Ecuación 1:

$$L = \frac{\sum_{j=1}^n w_j \log(1 + \exp(-2y_j f(x_j)))}{\sum_{j=1}^n w_j}, \quad (1)$$

En donde “ w_j ” es el peso para la observación “ j ”, del software se normaliza de los pesos para sumar a 1, “ w_j ” = $\{-1,1\}$; y $f(w_j)$ es el resultado de la observación “ j ”. (ver Anexo: *Tablas comparativas del error de entrenamiento*), donde son mostrados una serie de experimentos en los cuales se hace una prueba de cada método para encontrar el hiperplano de separación usando cada función kernel para mapear los datos de entrenamiento como se muestra en la Figura 31.

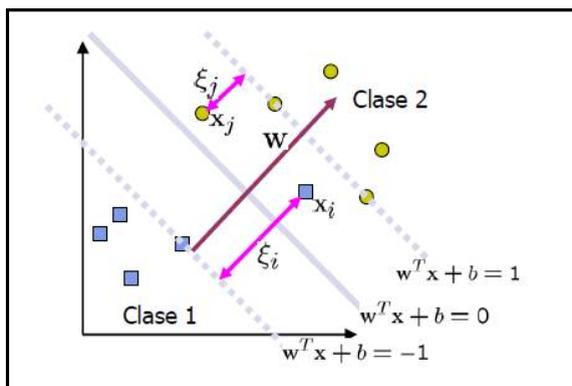


Figura 31. Error ξ_i de clasificación. Fuente: Autores.

Estas pruebas muestran los resultados obtenidos al combinar cada una de las funciones y kernel del algoritmo de clasificación

“SVM”, lo cual ayuda a elegir la función y el kernel que generan menos error, esto implica tener mayor eficiencia al momento de realizar el entrenamiento del algoritmo de inteligencia artificial a usar. El error de clasificación es una medida de error de clasificación binaria o de dos clases que tiene la forma mostrada en la Ecuación 2.

$$L = \frac{\sum_{j=1}^n w_j e_j}{\sum_{j=1}^n w_j}, \quad (2)$$

En donde “ w_j ” es el peso para la observación “ j ”, del software normalizado de los pesos para sumar a 1; $e_j = 1$, si la clase predicha de la observación “ j ” difiere de su verdadera clase y 0 en caso contrario. Por otro lado, la puntuación “SVM” para clasificar la observación “ x ” es la distancia de separación del hiperplano de clasificación, en donde el signo de esta observación en la frontera de decisión está entre $-\infty$ y $+\infty$. Así, una puntuación positiva para una clase indica que “ x ” tiene una probabilidad más alta de estar en esa clase. Por otro lado, una puntuación negativa indica lo contrario, Ecuación 3.

$$f(x) = \sum_{j=1}^n \alpha_j y_j G(x_j, x) + b, \quad (3)$$

Donde $(\alpha_1, \dots, \alpha_n, b)$ son los parámetros estimados de la “SVM”, $G(x_j, x)$ es el producto escalar en el espacio predictor entre “ x ” y los vectores de soporte, y la suma incluye las observaciones conjunto de entrenamiento. Si $G(x_j, x) = x_j'x$ (el núcleo lineal), entonces la puntuación de la función se reduce en la Ecuación 4 a:

$$f(x) = (x/s)' \beta + b, \quad (4)$$

Se tiene que “ s ” es la escala del núcleo y “ β ” es el vector de coeficientes lineales. El diagrama de bloques de los comandos del software para SVM para obtener el modelo de entrenamiento se muestra en la Figura 32.

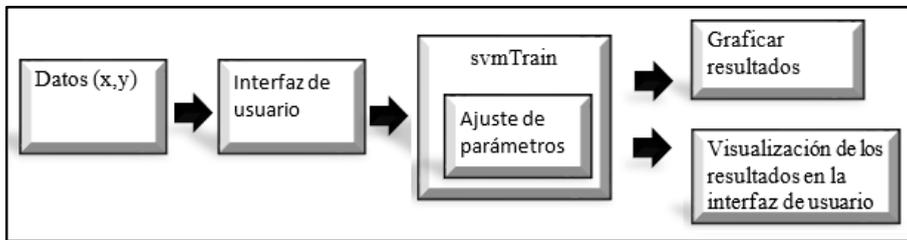


Figura 32. Comandos para obtener el modelo de entrenamiento.
Fuente MSV LS Toolbox Matlab

Las redes neuronales artificiales (RNA) son redes interconectadas masivamente en paralelo de elementos simples (*usualmente adaptativos*) y con organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema nervioso biológico. Debido a su fundamentación biológica, las redes neuronales artificiales presentan un gran número de características semejantes a las del cerebro, por ejemplo, son capaces de aprender de la experiencia, de generalizar de casos anteriores a nuevos casos, de abstraer características esenciales a partir de entradas que representan información irrelevante, entre otros ejemplos. Debido a esto, las redes neuronales artificiales ofrecen numerosas ventajas y son aplicadas en múltiples áreas. Dentro de las ventajas que las redes neuronales pueden ofrecer están: *i) aprendizaje adaptativo*: capacidad de aprender a realizar tareas basadas en un entrenamiento o en una experiencia inicial; *ii) auto-organización*: una red neuronal puede crear su propia organización o representación de la información que recibe mediante una etapa de aprendizaje; *iii) tolerancia a fallos*: la destrucción parcial de una red conduce a una degradación de su estructura; sin embargo, algunas capacidades de la red se pueden retener, incluso sufriendo un gran daño; *iv) operación en tiempo real*: los cómputos neuronales pueden ser realizados en paralelo; para esto se diseñan y fabrican máquinas con hardware especial para obtener esta capacidad; y *v) fácil inserción en la tecnología existente*: se pueden obtener chips especializados para redes neuronales que mejoran su capacidad en ciertas tareas. Ello facilitará la integración modular en los sistemas existentes.

La distribución de neuronas dentro de la red se realiza formando niveles o capas, con un número determinado de dichas neuronas en cada una de ellas. Un ejemplo se muestra en la Figura 33.

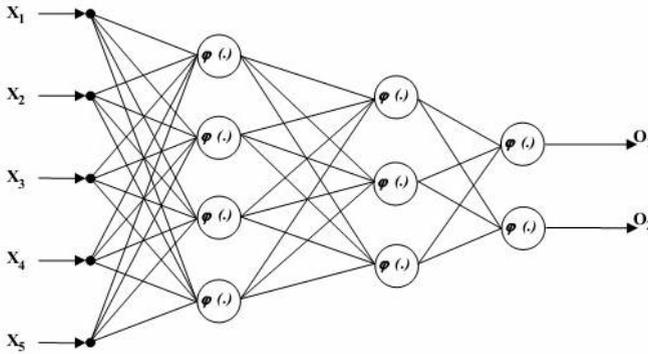


Figura 33. Esquema de una red neuronal. Fuente Scielo. Fuente: Autores.

A partir de su situación dentro de la red, se pueden distinguir tres tipos de capas: i) *capas de entrada*: es la capa que recibe directamente la información proveniente de las fuentes externas de la red; ii) *capas ocultas*: son internas a la red y no tienen contacto directo con el entorno exterior. El número de niveles ocultos puede estar entre cero y un número elevado. Las neuronas de las capas ocultas pueden estar interconectadas de distintas maneras, lo que determina, junto con su número, las distintas topologías de redes neuronales; y iii) *capas de salidas*: transfieren información de la red hacia el exterior.

En la *Figura 19*, se puede observar el ejemplo de la estructura de una posible red multicapa, en la que cada nodo o neurona únicamente está conectada con neuronas de un nivel superior. Se puede resaltar que hay más conexiones que neuronas en sí; en este sentido, se dice que una red es totalmente conectada si todas las salidas desde un nivel llegan a todos y cada uno de los nodos del nivel siguiente. En 1986, Rumelhart, Hinton y Williams, formalizaron un método para que una red neuronal aprendiera la asociación que existe entre los patrones de entrada y las clases correspondientes, utilizando varios niveles de neuronas. El algoritmo de aprendizaje supervisado usado para entrenar redes neuronales artificiales escrito se denomina

“Red neuronal *backpropagation*”, esto es, la propagación hacia atrás de errores o retropropagación (del inglés *backpropagation*), está en la capacidad que tiene de auto adaptar los pesos de las neuronas de las capas intermedias para aprender la relación que existe ente un conjunto de patrones de entrada y sus salidas correspondientes. Es importante resaltar la capacidad de generalización, o sea, la facilidad de dar salidas satisfactorias a entradas que el sistema no ha visto nunca en su fase de entrenamiento. La red debe encontrar una representación interna que le permita generar las salidas deseadas cuando se le dan entradas de entrenamiento, y que pueda aplicar, además, a entradas no presentadas durante la etapa de aprendizaje para clasificarlas.

Para entrenar una red neural de tipo “*backpropagation*” se debe seguir los siguientes pasos: I) Inicializar los pesos de la red con valores pequeños aleatorios, II) Presentar un patrón de entrada y especificar la salida deseada que debe generar la red; III) Calcular la salida actual de la red. Para ello se presentan las entradas a la red y se calcula la salida de cada capa hasta llegar a la capa de salida, ésta será la salida de la red. A su vez, para el paso tres se deben realizar los siguientes tres pasos: i) Se calculan las entradas netas para las neuronas ocultas procedentes de las neuronas de entrada. Para una neurona “j” oculta, Ecuación 5.

$$\text{net}_{pj}^h = \sum_{i=1}^N w_{ji}^h X_{pi} + \theta_j^h \quad , \quad (5)$$

En donde el índice “h” se refiere a magnitudes de la capa oculta; el subíndice “p”, al p-ésimo vector de entrenamiento, y “j” a la j-ésima neurona oculta. El término “ θ ” puede ser opcional, pues actúa como una entrada más. ii) Se calculan las salidas de las neuronas ocultas, Ecuación 6:

$$y_{pj} = f_j^h(\text{net}_{pj}^h) \quad , \quad (6)$$

iii) Se realizan los mismos cálculos para obtener las salidas de las neuronas de salida, Ecuación 7 y Ecuación 8

$$\text{net}_{pk}^o = \sum_{j=1}^L w_{kj}^o y_{pj} + \theta_k^o \quad , \quad (7) \quad y_{pk} = f_k^o(\text{net}_{pk}^o) \quad , \quad (8)$$

IV) Calcular los términos de error para todas las neuronas. Si la neurona k es una neurona de la capa de salida, el valor de la delta es se muestra Ecuación 9.

$$\delta_{pk}^o = (d_{pk} - y_{pk}) f_k^{o'}(\text{net}_{pk}^o) \quad , \quad (9)$$

La función f debe ser derivable. En general disponemos de dos formas de función de salida, Ecuación 10 y Ecuación 11.

a) La función lineal:

$$f_k(\text{net}_{jk}) = \text{net}_{jk} \quad , \quad (10)$$

b) La función sigmoideal:

$$f_k(\text{net}_{jk}) = \frac{1}{1 + e^{-\text{net}_{jk}}} \quad , \quad (11)$$

La selección de la función depende de la forma que se decida representar la salida: si se desea que las neuronas de salida sean binarias, se utiliza la función sigmoideal, en otros casos, la lineal. Para una función lineal, se tiene, Ecuación 12.

$$f_k^{o'} = 1 \quad , \quad (12)$$

Mientras que la derivada de una función sigmoideal es, ecuación 13.

$$f_k^{\circ'} = f_k^{\circ}(1 - f_k^{\circ}) = y_{pk}(1 - y_{pk}) \quad , \quad (13)$$

Por lo que los términos de error para las neuronas de salida quedan, ecuación.

a) Para la salida lineal: $\delta_{pk}^{\circ} = (d_{pk} - y_{pk})$

b) Para la salida sigmoideal: $\delta_{pk}^{\circ} = (d_{pk} - y_{pk})y_{pk}(1 - y_{pk})$

Si la neurona “j” no es de salida, entonces la derivada parcial del error no puede ser evaluada directamente, por tanto se obtiene el desarrollo a partir de valores que son conocidos y otros que pueden ser evaluados. La expresión obtenida en este caso es, Ecuación 14:

$$\delta_{pj}^h = f_j^{h'}(\text{net}_{pj}^h) \sum_k (\delta_{pk}^{\circ} w_{kj}^{\circ}) \quad , \quad (14)$$

Donde se observa que el error en las capas ocultas depende de todos los términos de error de la capa de salida. De aquí surge el término propagación hacia atrás; V) Actualización de los pesos: para se utiliza un algoritmo recursivo, comenzando por las neuronas de salida y trabajando hacia atrás hasta llegar a la capa de entrada, ajustando los pesos de la siguiente forma:

a) Para los pesos de las neuronas de la capa de salida, Ecuación 15 y Ecuación 16

$$w_{kj}^{\circ}(t + 1) = w_{kj}^{\circ}(t) + \Delta w_{kj}^{\circ}(t + 1) \quad , \quad (15)$$

$$\Delta w_{kj}^{\circ}(t + 1) = \alpha \delta_{pk}^{\circ} y_{pj} \quad , \quad (16)$$

b) Para los pesos de las neuronas de la capa oculta, Ecuación 17 y Ecuación 18.

$$w_{ji}^h(t+1) = w_{ji}^h(t) + \Delta w_{ji}^h(t+1) \quad , \quad (17)$$

$$\Delta w_{ji}^h(t+1) = \alpha \delta_{pj}^h x_{pi} \quad , \quad (18)$$

En ambos casos, para acelerar el proceso de aprendizaje se puede añadir un término momento; Y VI) El proceso se repite hasta que el término de error resulta aceptablemente pequeño para cada uno de los patrones aprendidos. El término de error es dado en la Ecuación 19.

$$E_p = \frac{1}{2} \sum_{k=1}^M \delta_{pk}^2 \quad , \quad (19)$$

El cálculo de estas medidas son necesarias en el momento realizar la clasificación ya que estas medidas permiten predecir con certeza si se presenta alguna de las clases usadas al momento de realizar la clasificación o simplemente si no se presenta, eso es mostrado en diagrama de bloques clasificación de la red neuronal mostrado en la Figura 34.

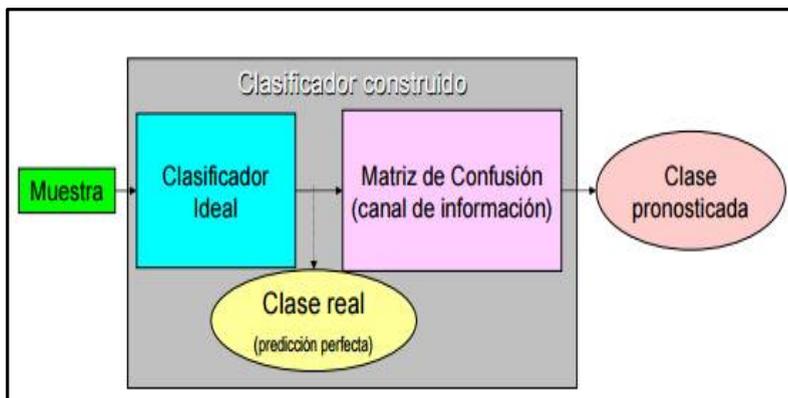


Figura 34. Diagrama de bloques clasificación de la red neuronal. Fuente: Autores.

Una vez realizado el entrenamiento de la base de datos utilizando los algoritmos de inteligencia artificial “SVM” y “RNA” los cuales se utilizan como un modelo clasificador, se hace necesario evaluar su desempeño como tal, en problemas de clasificación supervisada. Para ello se utilizan criterios tales como: porcentaje de clasificaciones correctas, diferentes medidas de error, medida F, entre otras. La capacidad del modelo para representar confiablemente el sistema real, se relaciona esencialmente con la precisión. No existe un modelo clasificador mejor que otro de manera general. Así, para cada problema nuevo es necesario determinar con cuál clasificador se pueden obtener mejores resultados, y es por esto que han surgido varias medidas para evaluar la clasificación y comparar los modelos empleados para un problema determinado. Las medidas más conocidas para evaluar la clasificación están basadas en la matriz de confusión que se obtiene cuando se prueba el clasificador en un conjunto de datos que no intervienen en el entrenamiento.

Tabla 31. Matriz de confusión

Matriz de Confusión		Clase verdadera	
		Pos	Neg
Clase Predicha pos	pos	VP	FP
neg	FN	VN	
Total columna		P	N

Fuente: Autores.

En la Tabla 31, se muestra la matriz de confusión de un problema de dos clases, donde Pos/pos es la clase positiva y Neg/neg la clase negativa, las siglas VP y VN representan los elementos bien clasificados de la clase positiva y negativa respectivamente, FP y FN identifican los elementos negativos y positivos mal clasificados respectivamente. Donde:

- VP: Corresponde al número de los datos de la enfermedad “Cirrosis” predichos correctamente.
- FN: Corresponde al número de los datos de la enfermedad “Hepatitis C” predichos incorrectamente.
- FP: Corresponde al número de los datos de la enfermedad “Cirrosis” predichos incorrectamente.
- VN: Corresponde al número de los datos de la enfermedad “Hepatitis C” predichos correctamente.
- P: Corresponde a la suma de los datos de la enfermedad “Cirrosis” predichos correctamente e incorrectamente.
- N: Corresponde a la suma de los datos de la enfermedad “Hepatitis C” predichos correctamente e incorrectamente.

En el aprendizaje supervisado, las métricas que se utilizan comúnmente para evaluar el rendimiento de generalización de los modelos inducidos, son el error de clasificación y la exactitud predictiva. En base a la matriz de confusión (Tabla 27), estas medidas se definen de en las Ecuaciones 20, 21 y 22:

$$\text{exactitud} = \frac{VP + VN}{P + N} \quad (20)$$

$$\text{Error} = \frac{FN + FP}{VP + FN + FP + VN} \quad (21)$$

$$\text{especificidad} = \frac{VN}{N} \quad (22)$$

Sin embargo, estas métricas no son apropiadas cuando las probabilidades a priori de las clases muy diferentes, ya que no consideran los costos en las clasificaciones incorrectas y son sensibles al sesgo entre las clases. Debido a que estas medidas dependen de la distribución de los datos, en problemas de aprendizaje no balanceado se adoptan otras métricas de evaluación que permiten medir el rendi-

miento sobre cada una de las clases de manera independiente. Algunas de éstas son la sensibilidad (*recall*) y la precisión, las cuales se definen a partir de la matriz de confusión en las Ecuaciones 23 y 24:

$$\text{sensibilidad} = \frac{VP}{P} \quad , \quad (23)$$

$$\text{precision} = \frac{VP}{VP + FP} \quad , \quad (24)$$

Como antecedentes del algoritmo de inteligencia artificial en aplicaciones biomédicas, se pueden citar a algunos investigadores se han dedicado al estudio de los algoritmos de inteligencia artificial tanto en el estudio biomédico como en la implementación de mini-computadores. Jorge E. Hernández L. y Santiago Salazar G. (2006), presentan el diseño y la implementación de una máquina de soporte vectorial “SVM” sobre un dispositivo lógico programable “PLD”. La máquina es utilizada para solucionar el problema fue la XOR, la cual implementaron sobre la tarjeta FPGA (ALTERA® FLEX10K). Más adelante Marcelo M. Raponi y Rodolfo O. Bonnin (2011), plantearon la realización de un dispositivo de rehabilitación visual basado en sistemas embebidos del tipo ARM. Este proyecto lo realizaron utilizando la mini computadora Beaglebone Black, para adquirir y procesar en tiempo real, señales de video provenientes de una mini cámara portátil, y generar un patrón de estimulación visual que brinde un cierto grado de rehabilitación a personas con baja visión. Por su parte, Luis Antonio Martín Nuez (2013-2014), realizó el diseño e implementación de una plataforma sensorial biométrica “e-Health Sensor Platform”, para el diagnóstico de pacientes en tiempo real y otras aplicaciones médicas. También Ch Srikanth, D.S. Pradeep M., Sreeram Charan K. (2014)”, cuyo proyecto tuvo un gran avance en la biomedicina, realizaron un dispositivo de apoyo en el diagnóstico médico que permite a los usuarios realizar un monitoreo del cuerpo utilizando las tarjetas Arduino y Beaglebone black y nueve sensores. La diferente información obtenida por estos sensores se pueden utilizar para controlar

en tiempo real el estado de un paciente o de obtener datos confidenciales con el fin de posteriormente analizarla para el diagnóstico médico. Por último, se menciona a Juan Pedro Domínguez et al. (2014) que plantearon realizar el “aprendizaje progresivo” basado en proyectos en el ámbito de la ingeniería biomédica en el diseño, construcción y programación de un ECG.

El tipo de aprendizaje utilizado para realizar la clasificación fue: I) Aprendizaje Supervisado, utilizado cuando se necesita un conjunto de datos de entrada previamente clasificado o cuya respuesta objetivo es conocida. Al momento de realizar el aprendizaje supervisado es necesario tener en cuenta las siguientes premisas: los datos incluyen características y etiquetas, el objetivo consiste en construir un estimador con capacidad predictiva sobre la etiqueta en función de un conjunto de características, los algoritmos supervisados se dividen a su vez en dos categorías: impacientes y perezosos; y en dos tipos de problemas: clasificación y regresión y en clasificación la etiqueta es discreta o categórica y en regresión es continua. Los usos de este tipo de aprendizaje son: *predicción de futuros casos*: se usa el modelo para predecir casos futuros; *extracción de conocimiento*: el modelo es más fácil de comprender (visualizar), retener y comprender; *comprensión*: el modelo es más simple que el conjunto de datos que explica; *detección de valores atípicos o anómalos*: excepciones que no son cubiertas por el modelo, por ejemplo fraudes o intrusiones. Y II) Aprendizaje off-line, que constituye una estrategia de aprendizaje en la cual se puede distinguir entre una fase de aprendizaje o una de entrenamiento, y una fase de operación o funcionamiento, existiendo un conjunto de datos de entrenamiento y un conjunto de datos de prueba que serán utilizados en la correspondiente fase. En otras palabras esta estrategia me permite realizar el entrenamiento de la *máquina de soporte vectorial “SVM”*, guardar el resultado (los vectores de soporte), y llevarlo a otro dispositivo diferente en el cual se puede probar los vectores de soporte ingresando algún dato desconocido para la “SVM” y ver el resultado obtenido.

4.4. Diseño de la aplicación como apoya al diagnóstico

En este apartado será discutida la fase de diseño del software realizada en la investigación hecha por Chacón y Miranda en el 2015, contemplando la integración que debe tener la aplicación de varias plataformas Investigación. A su vez, serán planteados y analizados los requerimientos funcionales y no funcionales en el momento de realizar la aplicación tomando como base la metodología SCRUM. Para realizar estos análisis, serán utilizados diagramas UML (lenguaje visual UML) con la finalidad de identificar más fácilmente los componentes principales del sistema, los usuarios y la interacción entre ellos (sistema-usuario). Además con estos diagramas puede ser observado el flujo de ejecución de tareas e trabajo para determinar si es fiable y consistente.

Los requerimientos funcionales y no funcionales en el momento de realizar la aplicación tomando como base la “Metodología SCRUM” se muestra en la Figura 35.

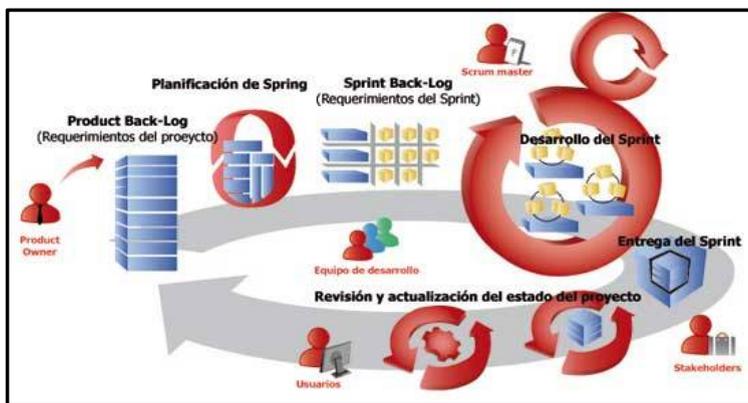


Figura 35. Metodología SCRUM para desarrollo de software - Aplicaciones complejas. Fuente: Autores.

Los requerimientos funcionales y no funcionales en el momento de realizar la aplicación tomando como base la metodología SCRUM se ilustran en la Tabla 32.

Tabla 32. Tabla de requerimientos

Tabla de requerimientos	
Funcionales	No funcionales
Recepción de exámenes	Interfaz grafica
Ingreso de datos	
Realizar diagnostico	
Validar datos	

Fuente: Autores.

De la tabla de requerimientos se identificaron los siguientes actores: a) paciente y b) especialista de la salud. del actor paciente se identificaron los siguientes casos de uso: a) entrega de exámenes y b) recepción de exámenes.

Del actor especialista de la salud se identificaron los siguientes casos de uso:

- Recepción de exámenes
- Ingreso de datos
- Realizar diagnóstico
- Entrega de diagnóstico

A partir del diagrama UML se identificaron los casos de usos y se generó la Figura 36.

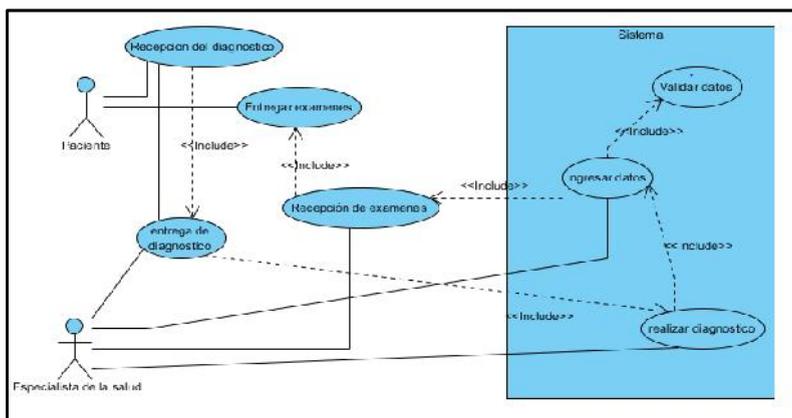


Figura 36. Diagrama de casos de usos. Fuente: Autores.

A partir de la identificación y análisis de los casos de uso se realizó en cada una de las iteraciones de construcción del software siguiendo la metodología SCRUM, esto es, generando el modelado UML de casos de uso que fue explicado anteriormente. Partiendo de las relaciones de generalización, composición, asociación o dependencia fueron construidas cada uno de los diagramas de secuencia UML el cual describe la interacción entre los usuarios o actores como se ilustra en la Figura 37.

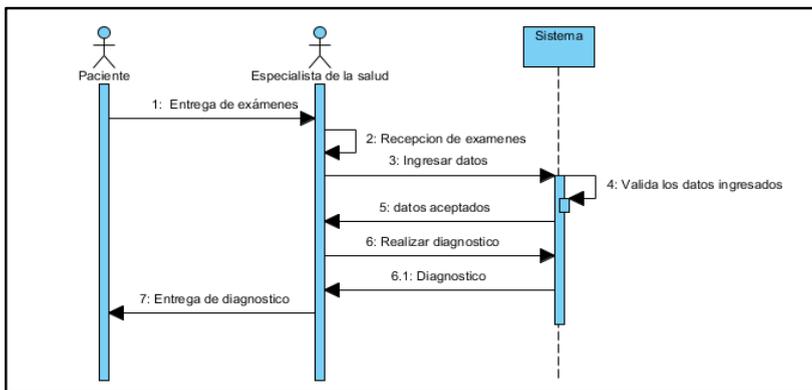


Figura 37. Diagrama de secuencia. Fuente: Autores.

A partir de los diagramas UML y siguiendo la metodología SCRUM se definieron dos tipos de requerimientos en los cuales los requerimientos funcionales son los que tienen mayor importancia y son esenciales para el funcionamiento de la aplicación; y los no funcionales son requerimientos los cuales no afectan el funcionamiento de la aplicación.

El desarrollo de la aplicación se dividió en las seis etapas siguientes: 1) selección del hardware utilizado; 2) elección del sistema operativo; 3) pre-procesamiento y tratamiento de los datos; 4) utilización del lenguaje unificado de modelado UML, el cual es un lenguaje gráfico para visualizar, especificar, construir y documentar un sistema; 5) realización del entrenamiento de la BD con aprendizaje supervisado ejecutado por un algoritmo “RNA” y “SVM” utilizando la biblioteca pybrain y sklearn; y 6) integración e implementación de la aplicación y el entrenamiento supervisado sobre la tarjeta de desarrollo BeagleBone Black.

La primera etapa se rigió bajo las normativas de desarrollo de software usando la metodología SCRUM. En las Tablas mostradas a continuación, especificadas cada una de ellas al final de la misma, son ilustradas las fases del inicio de aplicación y las tareas a realizar durante cada sprint, en donde un proyecto se ejecuta en bloques temporales cortos y fijos (iteraciones de un mes natural y hasta de dos semanas). Al finalizar cada sprint, se realiza una reunión en donde se analiza a través de la creación de una tabla que muestra si las tareas fueron realizadas o si aún siguen en proceso de construcción. Así, cada tarea que no se realice durante un sprint, pasa al siguiente sprint con mayor prioridad que las mismas tareas que se han propuesto para el siguiente sprint, las tareas que se van a realizar en el siguiente sprint son propuestas por el SCRUM manager, y así sucesivamente se van realizando nuevos sprint hasta que la aplicación cumpla con los requerimientos.

Se utilizó lenguaje C++ para calcular, implementar y mostrar los resultados de las SVM y Python ninja-IDE libre, para implementar, calcular y mostrar los resultados del algoritmo de la RNA.

4.5. Algoritmo de clasificación inteligente para el diagnóstico

Para la creación de un algoritmo de clasificación inteligente para el diagnóstico biomédico se comenzó inicialmente realizando pruebas que consistían en usar parte de la data hepática que no fue utilizada en el entrenamiento, a esto se le conoce con el nombre de pruebas supervisadas. Este proceso es ilustrado en la Figura 38.

Los resultados de las pruebas supervisadas en las “SVM” desarrollada en C ++, de las pruebas supervisadas “RNA” desarrollada en Python y las pruebas supervisadas en las “SVM” desarrollada en Python respectivamente se ilustran en las tablas adjuntas, en donde cada tabla tiene la información del total de datos utilizados para cada prueba, el valor promedio que tiene el tiempo al ingresar los datos, así como, la predicción correcta o incorrecta en cada clase (Tablas 33, 34, 35 y 36).



Figura 38. Muestra datos de entrenamiento y datos de prueba. Fuente: Autores.

Tabla 33. Datos de pruebas supervisadas SVM C++

Resultados de pruebas supervisadas SVM hecha en C++			
Total datos usados	60		
Tiempo promedio de los datos	El tiempo se calcula desde el momento de empezar a insertar el primer dato en la aplicación 2 minutos		
Predichos correctamente	Clase 1 24	Clase 2 18	Clase falsa 18
Predichos incorrectamente	Clase 1 0	Clase 2 0	Clase falsa 0

Fuente: Autores.

Tabla 34. Datos de pruebas supervisadas “RNA” Python

Resultados de pruebas supervisadas RNA hecha en Python			
Total datos usados	60		
Tiempo promedio de los datos	El tiempo se calcula desde el momento de empezar a insertar el primer dato en la aplicación 2 minutos		
Predichos correctamente	Clase 1 10	Clase 2 7	Clase falsa 9
Predichos incorrectamente	Clase 1 14	Clase 2 11	Clase falsa 9

Fuente: Autores.

Tabla 35. Datos de pruebas supervisadas “SVM” con procesamiento paralelo Python

Resultados de pruebas supervisadas SVM con procesamiento en paralelo hecha en Python			
Total datos usados	60		
Tiempo promedio de los datos	El tiempo se calcula desde el momento de empezar a insertar el primer dato en la aplicación 2 minutos		
Predichos correctamente	Clase 1 20	Clase 2 16	Clase falsa 17
Predichos incorrectamente	Clase 1 4	Clase 2 2	Clase falsa 1

Fuente: Autores.

Tabla 36. Datos de pruebas supervisadas “SVM” con procesamiento en cascada Python

Resultados de pruebas supervisadas SVM con procesamiento en cascada hecha en Python			
Total datos usados	60		
Tiempo promedio de los datos	El tiempo se calcula desde el momento de empezar a insertar el primer dato en la aplicación 2 minutos		
Predichos correctamente	Clase 1 24	Clase 2 18	Clase falsa 18
Predichos incorrectamente	Clase 1 0	Clase 2 0	Clase falsa 0

Fuente: Autores.

4.6. Análisis de los resultados de las pruebas supervisadas

En la Figura 39 se muestra el análisis comparativo hecho a las pruebas supervisadas de las Tablas anteriores. Este análisis es basado en dos puntos, los cuales son los más representativos en el ámbito de la medicina, “especificidad y sensibilidad”, estos valores fueron calculados utilizando las expresiones de las Ecuaciones 22 y 23.

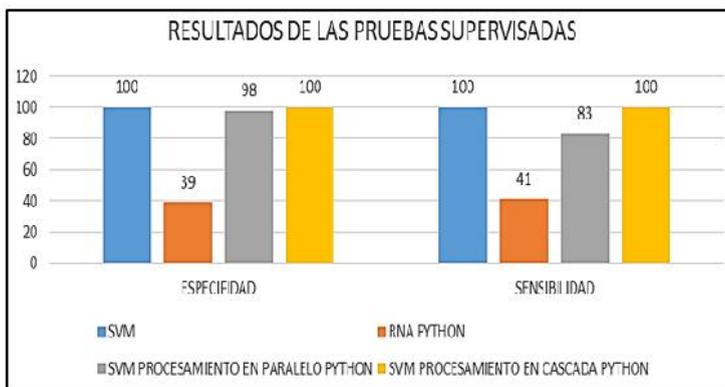


Figura 39. Comparación entre los resultados de las pruebas supervisadas.

En la gráfica se muestra que en este proyecto el algoritmo de clasificación “SVM” fue superior a “RNA”. Los resultados de las pruebas del entrenamiento de los algoritmos de inteligencia artificial dan como resultado que se obtiene una mayor efectividad de clasificación con el algoritmo de las “SVM” programado en C++ y Python en comparación con el algoritmo de la “RNA” programado en Python.

Las pruebas no supervisadas las cuales fueron realizadas por el Médico Cirujano Miguel Ángel Sanabria Ardila, identificado con la cedula de ciudadanía 80092649 de Bogotá. Con estudios de especialista en “gerencia y auditoría” y estudios de maestría en “efectividad clínica”; el cual realizó una evaluación de los datos utilizados en el proceso de entrenamiento del algoritmo de clasificación. En la Tabla 37 se ve reflejado el tamaño de los datos y el tiempo de demora al ser realizada la prueba por el especialista de la salud.

Tabla 37. Datos de pruebas realizadas por el especialista de la salud

Resultados de pruebas realizadas por el especialista de la salud	
Total datos usados	60
Tiempo promedio de los datos	El tiempo se calcula desde el momento de empezar a insertar el primer dato en la aplicación 2 minutos

Fuente: Autores.

Luego de haber realizado las pruebas a través de la aplicación para realizar diagnóstico de enfermedades hepáticas utilizadas en este proyecto *“Colelitiasis y cirrosis hepática ligera”* por el especialista de la salud, se recomendó realizar una comparación con dos pruebas tradicionales más representativas, la prueba de *“gold standard”* y la prueba *“ecográfica”* (Figura 40).

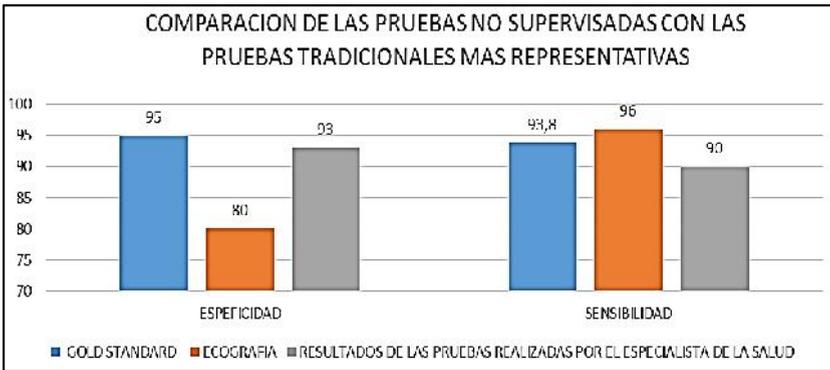


Figura 40. Comparación de pruebas no supervisadas y pruebas tradicionales.
Fuente: Autores.

Luego de realizar las pruebas el especialista de la salud realizó los siguientes comentarios y conclusiones sobre el proyecto y los resultados obtenidos: *“El desarrollo de herramientas informáticas para el soporte de las decisiones clínicas tiene especial relevancia en los sistemas de salud del mundo que enfrentan retos de atención a sus poblaciones, en especial en el ejercicio clínico individual del día a día, donde los médicos enfrentamos barreras de tiempo, conocimiento y*

costos, por esta razón el desarrollo de aplicaciones que ahorren costos y/o dinero a los sistemas de salud tiene especial importancia en cualquier país del mundo. Este trabajo es una apuesta interesante por desarrollar una aproximación paraclínica al diagnóstico de la enfermedad hepática, que hoy por hoy sigue siendo un reto clínico”. “La predicción en medicina es un deseo generalizado de los médicos clínicos y afinar las herramientas que ayudan al diagnóstico mediante métodos estadísticos tiene un valor que trasciende a la seguridad de los pacientes, y por este solo hecho todos los estudios y proyectos que se desarrollen sin importar en principio los resultados, aportan a esta loable labor. Este trabajo presenta en Colombia un interesante inicio de la aproximación paraclínica a través de las tecnologías que sin duda a medida que se vaya refinando en datos estadísticos y eventual reducción en el número de variables, pueda tanto ayudar a mejorar la certeza diagnóstica (sensibilidad y especificidad) como a aliviar los costos de exámenes innecesarios al sistema de salud. Las pruebas realizadas con las bases de datos arrojan resultados coherentes con las patologías estudiadas, no obstante es necesario tener en cuenta que la interpretación de las mismas debe asumirse necesariamente en el contexto individual de cada paciente, es decir en primera instancia no podría esperarse una certeza diagnóstica muy eficiente, toda vez que los pacientes con enfermedad hepática suelen convivir con otras enfermedades y variables psicosociales que naturalmente desbordan el alcance del presente proyecto. Pero es muy importante empezar por ayudar al médico en sus decisiones diarias con herramientas de este tipo”.

Las conclusiones obtenidas de la aplicación biomédica sobre una tarjeta de desarrollo de acuerdo a las pruebas realizadas por el especialista de la salud y comparándola con las pruebas tradicionales más representativas, arrojan resultados coherentes con las patologías estudiadas como se muestra en la gráfica de “Resultados de Pruebas no Supervisadas y Pruebas Tradicionales más Representativas”, por lo cual al analizar los resultados obtenidos durante todo el estudio de este trabajo de “investigación, desarrollo e innovación” y tomándose como soporte los consejos y conclusiones dadas por el especialista

de la salud, demuestran que la aplicación tiene resultados eficientes, y el proyecto ha de representar un gran inicio para la realización de nuevas herramientas de aplicaciones biomédicas de vanguardia, para los especialistas del amplio espectro de la salud.

Referencias

- Aracil, J. (2006). Ingeniería y pensamiento. Fundación El Monte. Sevilla. España.
- Arduino Projects Book (2015). Manual Oficial de Arduino. Traducido por Tino Fernández, Profesor de electrónica e informática at Teacher Training in FP.
- Balestrini, M. (2002). Cómo se Elabora el Proyecto de Investigación. BL Consultores, Caracas.
- Boon, M. (2009). Understanding in the Engineering Sciences: Interpretative Structures. In: Scientific Understanding: Philosophical Perspectives.
- Computadoras y Cosas Pequeñas (2013). Recuperado de: <http://cerebrosartificiales.blogspot.com>.
- Chacón R. José G. y Ramírez Carlos (Tutor) - (2005). Sistema Difuso Neuronal Aplicado a la Clasificación de Enfermedades por Patrones Solapados. UNET. Venezuela.
- Bucci, N. y Terán A. (2008). Nuevas responsabilidades de los ingenieros. Revista universidad, ciencia y tecnología. Volumen 12, N° 47. UNEXPO.
- Dias, A. (2008). Toward an Epistemology of Engineering. The Royal Academy of Engineering, London.
- Dias, A. (2014). De la Nature Historique Des Pratiques D'Ingénierie. Revue d'anthropologie des connaissances. Vol. 8, N° 2.
- Dresch, A.; Lacerda, D., y Miguel, P. (2015). A Distinctive Analysis of Case Study, Action Research and Design Science Research. Review of Business Management, São Paulo, Vol. 17, No. 56.
- Fernández, A., García, S. y Herrera, F. (2011). Addressing the Classification with Imbalanced Data: Open Problems and New Challenges on Class Distribution. E.

- García Córdoba, F. (2007). *La Investigación tecnológica*. 2ª Edición. Limusa, México.
- García Córdoba, F., Morales, A. y Muñoz, C. (2010). *Revista UPIICSA*, XVIII, VII, No. 53-54.
- Gregor, S. (2009). *Building Theory in the Sciences of the Artificial*. En *Proceedings of the 4th international conference on design science research in information systems and technology*. ACM.
- Hendricks, V., Jacobsen, A. y Pedersen, S. (2000). *Identification of matrices in science and engineering*. *Journal for General Philosophy of Science*, No. 31.
- Hilera José R. y Martínez Víctor J. (2000). *Redes neuronales artificiales*. Alfa omega-Rama.
- Hurtado de Barrera, J. (2010). *El proyecto de investigación*. Sexta Edición. Ediciones Quirón, Bogotá, Colombia.
- Hurtado de Barrera, J. (2012). *Metodología de la investigación: guía para una comprensión holística de la ciencia*. Cuarta Edición. Ediciones Quirón, Venezuela.
- Jaimes, Rosalvina (2012). *Origen y destino del conocimiento científico*. Fondo Editorial Tropikos, Caracas.
- Johannesson, P., Perjons, E., & Bider, I. (2013). "What are the siblings of design science research". In *SIG Prag Workshop on IT Artefact Design & Workpractice Improvement*.
- Johannesson, P. y Perjons, E. (2014). *An Introduction to Design Science*. Springer Cham, Suiza.
- Kirby, R., Withington, S., Darling, A. y Kilgour, F. (1990). *Engineering in History*. Dover Publications. USA.
- Larsen, M. (2005). "Enabling Innovation through Engineering: The Other Face of University Research". Department of Industrial Economics and Strategy Summer Conference 2005 on Dynamics of Industry and Innovation: Organizations, Networks and Systems.
- Love, T. (1998). *Social, environmental and ethical factors in engineering design theory: a post-positivist approach*. Tesis Doctoral. The University of Western Australia.

- Politécnico Santiago Mariño (2012). Manual de Líneas de Investigación de Ingeniería de Sistemas. Departamento de Investigación. Valencia, Estado Carabobo, Venezuela. 1era edición.
- Maldonado, C. y Gómez; N. (2012). The complexification of engineering. Complexity, vol. 17, no 4.
- Marjoram, T. y Zhong, Y. (2010). What engineering is, what engineers do. UNESCO. Engineering: Issues, Challenges and Opportunities for Development. Publishing, Francia.
- Martín Bonifacio y Sanz Alfredo (2002). Redes neuronales y sistemas difusos, Alfa omega-Rama.
- Méndez, C. (2011). Metodología. 4ª Edición. Limusa, México. Oxford Dictionaries (en línea). <http://www.oxforddictionaries.com/definition/english/engineering>.
- Peffer, K., Tuunanen, T., Rothenberger, M. A. y Chatterjee, S. (2008). A design science research methodology for information systems research.
- Phillips, E. y Pugh, D. (2008). La tesis doctoral. Bresca Editorial, Barcelona, España.
- Pomares B. Jorge (2009). Manual de arduino. Grupo de innovación Educativa en Automática. Universidad de Alicante. GITE - IEA.
- Politécnico Santiago Mariño (2012). Líneas de Investigación de la Carrera Ingeniería de Sistemas. Escuela de Ingeniería de Sistemas. Extensión Valencia. Venezuela.
- Ramírez, T. (2010). Cómo hacer un proyecto de investigación. Editorial Panapo, Caracas.
- Raspberry pi or beaglebone black (2014). <http://makezine.com/magazine/how-to-choose-the-right-platform-raspberry-pi-or-beaglebone-black/> How to choose the right platform.
- Revista Médica de Chile (2008). “Estudios que Evalúan un Test Diagnóstico: Interpretando sus Resultados”. Sociedad Médica de Santiago versión impresa ISSN 0034-9887
- Sabino, C. (1992). El proceso de investigación. Editorial Panapo, Caracas, Venezuela.
- Simon, H. (1996). The sciences of the artificial. Third Edition. MIT Press.

- Turrioni, J. y Mello, C. (2012). Metodologia de Pesquisa em Engenharia de Produção. UNIFEI, Itajubá, Brasil.
- UPEL (2011). Manual de trabajos de grado de maestría y tesis doctorales. 4ª Edición. FEDUPEL, Caracas, Venezuela.
- Vaishnavi, V. y Kuechler, W. (2015). Design Science Research Methods and Patterns: Innovating Information and Communication Technology. 2nd Edition. Boca Raton, FL.
- Van Rijsbergen, C. J. (1979). Information retrieval. London, Butterworths.
- Verschuren, Piet y Hartog, Rob (2005). Evaluation in design-oriented research. Quality y Quantity, Vol. 39.

Anexo

Tablas comparativas del error de entrenamiento

Tabla 1. Comparativa del error usando el método SMO

Tipo de Kernel	Tipo de enfermedad	
	Cirrosis hepática	Colelitiasis
Lineal	0.404255319148936	0.446808510638298
RBF	0.170212765957447	0.127659574468085
Cuadrático	0.191489361702128	0.446808510638298
Polinomial	0.127659574468085	0.361702127659575
Perceptron multicapa	0.148936170212766	0.255319148936170

Error de entrenamiento método SMO. Fuente Autor.

Tabla 2. Comparativa del error usando el método QP

Tipo de Kernel	Tipo de enfermedad	
	Cirrosis hepática	Colelitiasis
Lineal	0.170212765957447	0.255319148936170
RBF	0.191489361702128	0.106382978723404
Cuadrático	0.148936170212766	0.255319148936170
Polinomial	0.191489361702128	0.191489361702128
Perceptron multicapa	0.127659574468085	0.255319148936170

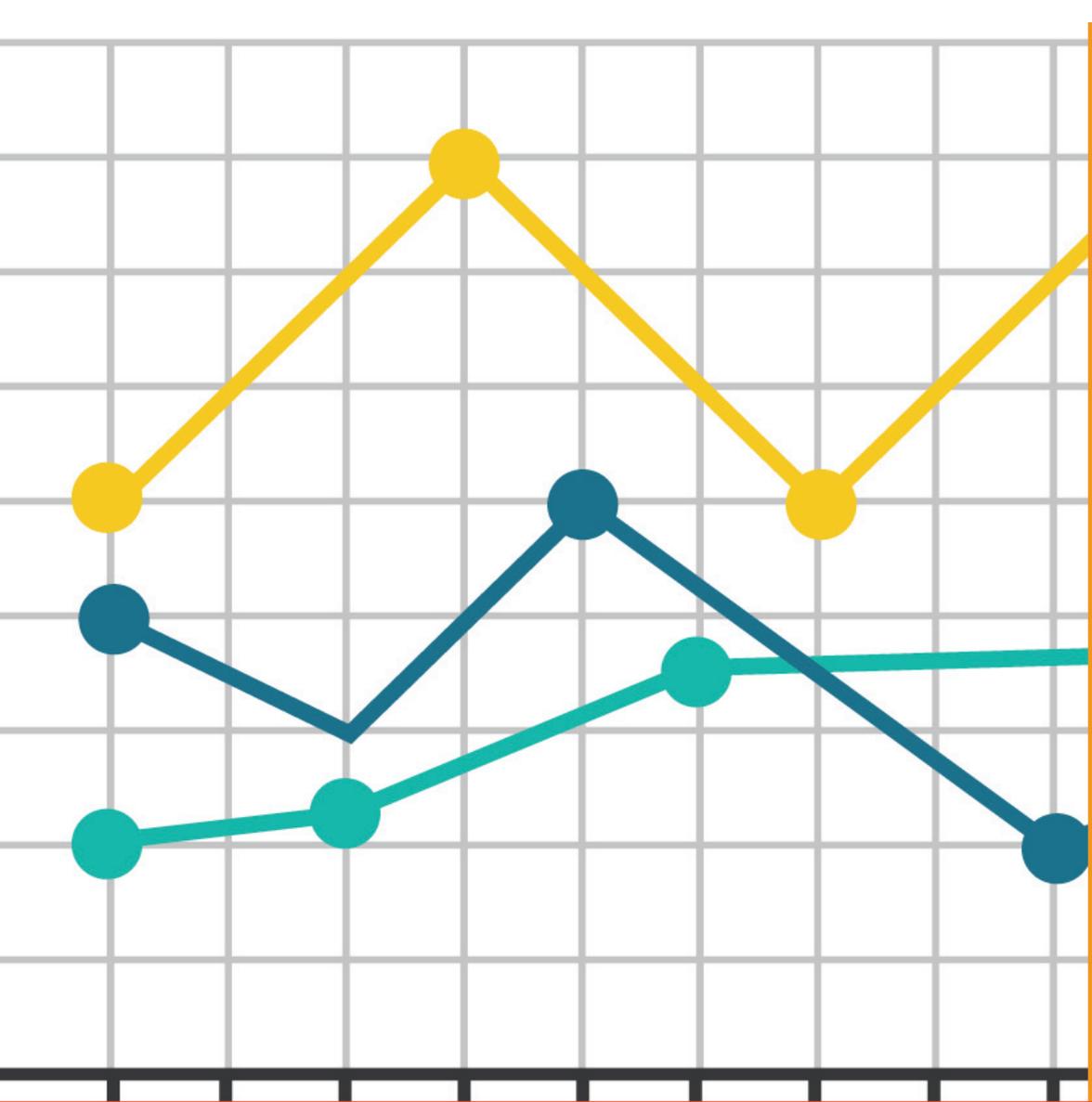
Error de entrenamiento método QP. Fuente Autor.

Tabla 3. Comparativa del error usando el método LS

Tipo de Kernel	Tipo de enfermedad	
	Cirrosis hepática	Colelitiasis
Lineal	0.276595744680851	0.212765957446809
RBF	0.148936170212766	0.212765957446809
Cuadrático	0.234042553191489	0.255319148936170
Polinomial	0.255319148936170	0.276595744680851
Perceptron Multicapa	0.319148936170213	0.276595744680851

Error de entrenamiento método MSV LS. Fuente Autor.





Libro electrónico