

Generalidades sobre Metodología de la Investigación

Dr. Manuel E. Cortés Cortés

Dra. Miriam Iglesias León

10
Colección Material Didáctico

DIRECTORIO GENERAL
Universidad Autónoma del Carmen

C. P. José Nicolás Novelo Nobles
Rector

Ing. Guadalupe de la Cruz Benítez
Secretario General

Mtro. Andrés Salazar Dzib
Secretario Académico

Dr. José Manuel Pérez Gutiérrez
Secretario de Extensión Universitaria

Psic. Juan Manuel Carrillo
Secretario de Servicios Estudiantiles

Lic. Francisco Javier Vera de la O.
Director de Planeación

Dr. Francisco Javier Ortega Quijano
Director de Investigación y Posgrado

Lic. Melenie Guzmán Ocampo
Directora de Difusión Cultural

Generalidades sobre Metodología de la Investigación

Dr. Manuel E. Cortés Cortés

Dra. Miriam Iglesias León



Universidad Autónoma del Carmen
Ciudad del Carmen, Campeche, México 2004

Primera edición, 2004

Generalidades sobre Metodología de la Investigación

Dr. Manuel E. Cortés Cortés

Dra. Miriam Iglesias León

D.R. © Universidad Autónoma del Carmen

Av. Concordia, calle 56, No. 4, CP. 24180

Ciudad del Carmen, Campeche, México

Tel: (01 938) 381 – 10 – 18

ISBN: 968 – 6624 – 87– 2

Encargada de la edición: Ana Polkey Gómez

Corrección de estilo: Eduardo Martínez Hernández

Ilustración de portada: Mario Brito Ávila

Impreso en México

“Investigar significa pagar la entrada por adelantado
Y entrar sin saber lo que se va a ver.”
Oppenheimer

“Es importante aprender método y técnicas de investigación,
Pero sin caer en un fetichismo metodológico.
Un método no es una receta mágica. Más bien
Es como una caja de herramientas, en la que se toma
Lo que sirve para cada caso y para cada momento.”
Ander-Egg

Capítulo 1: La Metodología de la Investigación Científica.

La Investigación Científica está encaminada a profundizar el conocimiento de un proceso ya sea teórico, práctico o teórico-práctico, parte del conocimiento científico y lo lleva a la solución de problemas de la sociedad que de una forma u otra no han sido investigados o su investigación se ha conducido en otra dirección.

La Investigación Científica surge de la necesidad del hombre de dar solución a los problemas mas acuciantes de la vida cotidiana, de conocer la naturaleza que lo rodea y transformarla en función de satisfacer sus intereses y necesidades. El carácter de la investigación científica es creativo e innovador aplicando lo último del conocimiento científico.

La Metodología es la ciencia que nos enseña a dirigir determinado proceso de manera eficiente y eficaz para alcanzar los resultados deseados y tiene como objetivo darnos la estrategia a seguir en el proceso.

La Metodología de la Investigación (M.I.) o Metodología de la Investigación Científica es aquella ciencia que provee al investigador de una serie de conceptos, principios y leyes que le permiten encauzar de un modo eficiente y tendiente a la excelencia el proceso de la investigación científica. El objeto de estudio de la M.I. Lo podemos definir como el proceso de Investigación Científica, el cual está conformado por toda una serie de pasos lógicamente estructurados y relacionados entre si. Este estudio se hace sobre la base de un conjunto de características y de sus relaciones y leyes.

En la Educación Superior la Investigación Científica tiene doble función: contribuye a la formación del profesional y es, además una vía para resolver los problemas que se presentan en la sociedad. Con el objetivo de formar profesionales con un alto grado de competencia y desempeño en las universidades latinoamericanas es necesario lograr que los procesos que en ella se desarrollan alcancen la excelencia académica y la excelencia científica.

La M. I. Estudia en fin de cuenta las características, las leyes y los métodos de dicho proceso, todo lo cual, en su conjunto constituye un modelo teórico de la Investigación Científica.

Algunos aspectos a tomar en cuenta para realizar una investigación:

- Surgimiento de la **IDEA**, el **TEMA** o el **AREA** que se desea investigar.
- Seleccionar el lugar dónde desarrollar el estudio.
- Elección de los participantes en el proceso de estudio.
- Revisión del lugar de estudio.
- Realización del trabajo de campo de la investigación.
- Diseñar la Investigación. Bosquejo del conjunto de los componentes (Tema, problema, objetivos, etc. es la dimensión estratégica del proceso de investigación.)
- Confección o selección del Instrumento.
- Etapa de la Recopilación de la Información.
- Procesamiento de la Información para su posterior análisis.
- Método del Análisis de Datos.
- Confección Final de los Resultados de la Investigación. Creación del Informe Final.

El **qué** de la Investigación Científica es el llamado **objeto de la investigación**, es aquella parte de la realidad objetiva sobre la cuál va a actuar el investigador (**sujeto**) desde el punto de vista, tanto práctico como teórico, para lograr la solución del problema. El objeto debe precisarse en el momento que se determina la investigación. Dentro del proceso de investigación el investigador va precisando el objeto y de aquí surge el **Campo de Acción**, el Campo de Acción es un concepto más estrecho que el Objeto pues es una parte de éste, una abstracción particular del primero donde el investigador va a trabajar mas directamente su investigación.

1.1. Los Enfoques de la Metodología de Investigación

Los enfoques más comunes en la M. I son el Enfoque Cualitativo y el Enfoque Cuantitativo, ambos desde su surgimiento han tenido diversidad de opiniones encontradas, diferencias sustanciales, críticas del uno al

otro, etc., pero si se puede establecer en forma general algunos puntos de contacto entre ambos:

- Se basan en observaciones y evaluaciones del fenómeno.
- Se llegan a conclusiones como resultados de esas observaciones y evaluaciones.
- De alguna forma, mas o menos fundamentada, demuestran el grado de realidad de las conclusiones arribadas.
- Comprueban las conclusiones arribadas y hasta son capaces de generar nuevas fundamentaciones, basándose en las tendencias encontradas.

En síntesis ambos enfoques pueden resumirse como:

- **Enfoque cualitativo:** es una vía de investigar sin mediciones numéricas, tomando encuestas, entrevistas, descripciones, puntos de vista de los investigadores, reconstrucciones los hechos, no tomando en general la prueba de hipótesis como algo necesario. Se llaman **holísticos** porque a su modo de ver las cosas las aprecian en su totalidad, como un TODO, sin reducirlos a sus partes integrantes. Con herramientas cualitativas intentan afinar las preguntas de investigación. En este enfoque se pueden desarrollar las preguntas de investigación a lo largo de todo el proceso, antes, durante y después. EL proceso es más dinámico mediante la interpretación de los hechos, su alcance es más bien el de entender las variables que intervienen en el proceso mas que medirlas y acotarlas. Este enfoque es más bien utilizado en procesos sociales.
- **Enfoque cuantitativo:** toma como centro de su proceso de investigación a las mediciones numéricas, utiliza la observación del proceso en forma de recolección de datos y los analiza para llegar a responder sus preguntas de investigación. Utiliza la recolección, la medición de parámetros, la obtención de frecuencias y estadígrafos de la población que investiga para llegar a probar las **Hipótesis** establecidas previamente. En este enfoque se utiliza necesariamente el Análisis Estadístico, se tiene la idea de investigación, las preguntas de investigación, se

formulan los objetivos, se derivan las hipótesis, se eligen las variables del proceso y mediante un proceso de cálculo se contrastan las hipótesis. Este enfoque es más bien utilizado en procesos que por su naturaleza puedan ser medibles o cuantificables.

Ambos enfoques tienen sus ventajas y desventajas, son más útiles en algunos u otros procesos y su aplicación si es verdaderamente aplicada y profundizada puede dar los resultados esperados al proceso investigativo. Hay detractores de ambos y críticas en su aplicabilidad, al enfoque cuantitativo se lo califica de mucho más impersonal, frío y limitado, mientras al enfoque cualitativo se lo critica por su subjetividad y ser muy especulativos. Ambos enfoques bien aplicados pueden llevar a su investigador a obtener los resultados deseados, sin embargo en la actualidad se está tendiendo a las investigaciones con **enfoques mixtos** en donde se puedan tener las bondades de cada uno de ellos por separado.

- **Enfoque Mixto:** En un **enfoque mixto** el investigador utiliza las técnicas de cada uno por separado, se hacen entrevistas, se realizan encuestas para saber las opiniones de cada cual sobre el tema en cuestión, se trazan lineamientos sobre las políticas a seguir según las personas que intervengan, etc., además esas encuestas pueden ser valoradas en escalas medibles y se hacen valoraciones numéricas de las mismas, se obtienen rangos de valores de las respuestas, se observan las tendencias obtenidas, las frecuencias, se hacen histogramas, se formulan hipótesis que se corroboran posteriormente. En este enfoque mixto se integran ambas concepciones y se combinan los procesos para llegar a resultados de una forma superior.

1.2: Pasos de la Investigación Científica:

El proyecto de la investigación nace de la Idea a Investigar, esta **Idea** inicialmente es muy vaga, ambigua, no presenta ninguna solidez y surge de la necesidad de resolver problemas de la vida cotidiana para no sólo

conocer la naturaleza sino transformarla en beneficio de sus intereses y los de la sociedad en su conjunto. La Idea de la Investigación surge de cualquier situación problémica que se base en alguna forma en el conocimiento científico, que se dirija al mejoramiento del conocimiento científico solución y que su solución se base en las principales categorías y regularidades del conocimiento científico del área de que se trate.

Una regla para la generación de buenas ideas parte de varias premisas: el investigador se siente motivado, excitado por la idea, estas ideas llevan en sí algo novedoso, que no es necesariamente nuevo, estas ideas pueden servir de base para nuevas teorías o soluciones prácticas del problema y con ellas pueden surgir nuevas dudas que a su vez lleven al hombre a nuevas ideas en otros campos de acción o a nuevas aplicaciones en el mismo campo de acción.

de una **Idea** de investigación puede surgir un **problema de investigación** si se cumplen las premisas siguientes:

- La idea debe conducir a un problema objetivo, es decir, responder a una necesidad de la sociedad, partir de un desconocimiento científico y dar como resultado la creación de un nuevo conocimiento. **Objetividad.**
- La idea debe ser precisa, no tener ambigüedades, debe estar bien claro el objetivo y las cuestiones particulares de interés. **Especificidad.**
- La idea debe conducir a un problema que sea soluble en un tiempo determinado, no puede llevar a algo rebuscado, insoluble o en extremo difícil de resolver, su forma de solución debe estar garantizada, la búsqueda de la información, los métodos de análisis de datos, los métodos de solución, etc. **Asequible.**

Planteamiento del problema de investigación:

Plantear el problema no es mas que **afinar** y **estructurar** mas formalmente la **idea de investigación**. El planteamiento del problema

conlleva en si las siguientes fases o aspectos a tener en cuenta en forma muy general.

1. Selección del Tema de Investigación. (Idea de investigación, Área en donde se va a aplicar,...)
2. Punto de Partida. (¿Es realmente de interés el tema, existe información del mismo, se sabe dónde se puede encontrar, se conocen en forma general los resultados que puede traer la investigación?)
3. Observación Directa. (Se debe conocer el objeto de investigación).
4. Consulta Bibliográfica. (Se debe hacer una búsqueda para documentarse en lo más que pueda sobre el material escrito referido a su investigación.)
5. Consulta con Expertos. (Los expertos pueden ayudar al investigador a delimitar el objeto de conocimiento.)
6. Definición del Tema y del título preliminar de la Investigación.

En forma mas específica el planteamiento del problema conlleva a tres aspectos fundamentales:

➤ **Objetivo de la Investigación:** ¿Qué pretende la investigación? Es el **para qué** de la investigación. En forma muy clara y precisa se deben formular los objetivos de la investigación que son las líneas directrices por la que se va a encaminar la investigación. El objetivo es la aspiración, el propósito, el para qué, se desarrolla la investigación y debe tener algunas de las siguientes características:

1. Debe ser orientador porque es el punto de referencia a partir del cual se va a encaminar todo
2. Debe expresarse en forma sintética y generalizadora.
3. Debe expresarse en un tono afirmativo.
4. Debe declararse en forma clara, precisa y sin ambigüedades.
5. Debe limitarse a los recursos con que se cuenta para desarrollarlos.
6. Debe ser posible de evaluar con lo cual se estaría evaluando la investigación.

El objetivo puede ser desglosado en Objetivo General o globalizador de la investigación y los Objetivos Específicos que son los propósitos específicos por los cuales se puede lograr el objetivo general. El objetivo al trabajar sobre el objeto de la investigación delimita el campo de acción.

La relación problema, objeto (campo de acción) y objetivos, es una relación que, con carácter de ley, se presenta en el proceso de Investigación Científica, que establece vínculos esenciales y obligatorios entre esas tres características.

- **Las Preguntas de Investigación.** Además de los objetivos concretos de la investigación, es conveniente plantear, el problema de investigación, a través de una o varias preguntas, que no son más que las interrogantes que tiene el investigador sobre lo que se pretende investigar o alcanzar. Las preguntas de investigación no siempre son utilizadas para plantear el problema en sí, pueden servir de apoyo al mismo complementando el problema y haciendo más claro lo que se pretende. Las preguntas pueden ser más o menos generales, pero en la mayoría de los casos es mejor que sean más precisas y que orienten hacia las respuestas que se buscan con la investigación.

Las preguntas de Investigación al igual que los objetivos pueden ser modificadas en el transcurso de la investigación o inclusive agregarse nuevas para cubrir de esta forma los diversos aspectos del problema de investigación.

- **La Justificación de la Investigación.** Debe justificarse la investigación en todos los casos, es necesario justificar el estudio exponiendo las razones que se tienen para ello. La Justificación de la Investigación significa **el por qué** de la investigación. La justificación de la investigación está en función de varias cuestiones:

1. La conveniencia. ¿Para qué sirve la Investigación ?
2. Relevancia social. ¿Cuál es la trascendencia para la sociedad ?

3. Implicaciones Prácticas. ¿Ayudará a resolver algún problema práctico?
4. Valor Teórico. ¿En el campo de la teoría sentará alguna pauta?
5. Utilidad. ¿Qué utilidad tendrá la solución de la investigación?

De lo antes expuesto podemos decir que la Investigación se justifica desde los siguientes puntos de vistas que pueden ser abordados:

1. Justificación Teórica. Razones que argumentan el deseo de verificar, rechazar o aportar aspectos teóricos referidos al objeto de conocimiento. Aquí se plantean las siguientes interrogantes: ¿Quieres ampliar un modelo teórico?, ¿Quieres contrastar la forma en que un modelo teórico se presenta en la realidad?, ¿Esperas que los resultados sean un complemento teórico de aquel que fundamentastes?
2. Justificación Metodológica: Razones que sustentan un aporte por la utilización o creación de instrumentos y modelos de investigación. ¿El resultado de la investigación dará una serie de pasos a seguir en investigaciones en esa línea?, ¿El resultado de la investigación es un instrumento, un modelo matemático o un software que pueda ser empleado en otras investigaciones?
3. Justificación Práctica: Razones que señalen que la investigación propuesta ayudará en la solución de problemas o en la toma de decisiones. ¿El resultado de la investigación tiene una aplicación concreta y puede mostrar resultados?, ¿El resultado de la investigación ayudará a mejorar o solucionar sistemas y o procedimientos de alguna empresa u organización?, ¿El resultado es una solución económica concreta, administrativa u otro resultado práctico diferente?

1.3. El marco teórico de la investigación

La Investigación a realizar debe tomar en cuenta el conocimiento previamente construido, pues esta forma parte de una estructura teórica ya existente. El Marco Teórico implica analizar teorías, investigaciones y

antecedentes que se consideren válidos para el encuadre del estudio pues la búsqueda y sistematización de aquellas teorías precedentes pueden ayudar en el análisis del problema a investigar. La elaboración del Marco Teórico se realiza mediante conceptos, magnitudes, variables, leyes y modelos que existen en la ciencia y que se sistematizan con el objetivo de determinar en qué medida estos contribuyen a la solución del problema investigado y en qué medida son insuficientes. EL Marco Teórico se inicia desde el mismo momento que se formula el problema, donde es necesario conocer sus antecedentes teóricos para precisar si el problema tiene carácter científico o no.

El investigador define y delimita según su criterio y de acuerdo a su marco teórico conceptos involucrados en las variables de investigación (marco conceptual).

Funciones del marco teórico

Algunos autores atribuyen diversas funciones al marco teórico:

1. Evita tomar caminos equivocados en el proceso de Investigación. Previene de errores que han sido cometidos en investigaciones anteriores.
2. Da ideas de cómo realizar la investigación, nos orienta al estudiar lo precedente.
3. Sirve de guía orientadora al investigador y lo aleja de caminos que no sean los fundamentales para lo que pretende realizar, busca el camino, concreta los límites del trabajo y permite obtener diversos puntos de vista.
4. Contribuye al establecimiento de un modelo teórico y de una hipótesis de trabajo. Conduce a las hipótesis que serán afirmadas o negadas posteriormente al hacer las pruebas pertinentes.
5. Da luz al estudio de nuevos problemas de investigación. Permite el conocimiento de nuevas líneas y áreas de investigación.
6. Da un marco de referencia para las futuras interpretaciones de los resultados obtenidos.

Tareas o etapas del marco teórico

La elaboración del marco teórico tiene dos **tareas o etapas** fundamentales en el proceso de la investigación:

1. **Revisión de la Literatura**, mediante la cual se consulta, extrae y recopila la información relevante sobre el problema a investigar.
2. **Sistematización de las teorías** existentes que posibilita determinar el grado que la misma explica el problema científico a investigar y el grado en que no, se adopta una teoría o se desarrolla una perspectiva teórica o de referencia.

Para comenzar con la revisión de la literatura, se recomienda la consulta de expertos y centros de información científica, debe seleccionarse la bibliografía más importante e iniciar un proceso de lectura que debe contribuir a la creación de las fichas y resúmenes de los autores más relevantes que tienen relación directa con el objeto de la investigación. Las fuentes literarias pueden ser de diferentes tipos:

1. **Fuentes primarias** que son los libros, artículos, revistas, monografías, tesis disertaciones, documentos oficiales, reportes de asociaciones, trabajos presentados en conferencias o seminarios, artículos periodísticos, testimonios de expertos, películas, documentales, videocintas, foros, páginas Web, artículos de Internet y otros.
2. **Fuentes secundarias** que son los resúmenes referencias en donde se mencionan y comentan brevemente artículos, libros, tesis, disertaciones y otros documentos, relevantes en el campo de dicha investigación.
3. **Fuentes terciarias** que son los compendios directorios de títulos, revistas, autores, organizaciones científicas, catálogo de libros y otros, son documentos que compendian nombres y títulos de estas fuentes antes mencionadas. Son útiles para detectar fuentes no documentales.

La recopilación de la bibliografía o de cualquier otra fuente debe ser cuidadosamente realizada en cuanto al registro de la información que ofrece. Algunos aspectos fundamentales a tener en cuenta son:

4. Título del trabajo.
5. Autor.
6. Año de publicación.
7. Editorial.
8. Lugar de la edición.
9. Número de páginas en el caso de libros o número de las páginas en el caso de artículos científicos.
10. Nombre del Sitio y dirección electrónica completa en el caso de páginas Web.
11. Otros aspectos de utilidad para la ubicación futura del material.

Conclusiones de la Revisión Bibliográfica

Como resultado de la de la revisión de la literatura se puede llegar a las siguientes conclusiones:

1. Existe una teoría científica, completamente desarrollada, capaz de describir o explicar el problema de investigación.
2. Existen varias teorías o generalizaciones empíricas que han sido demostradas y que son aplicables a nuestro problema de investigación.
3. Existen partes de teorías con apoyo empírico o de modelación que explican algunas de las variables básicas de nuestro problema de investigación.
4. Hay descubrimientos interesantes, pero parciales, sin llegar a ajustarse a una teoría.
5. Existen solo guías aún no estudiadas e ideas relacionadas con el problema de investigación.

En los casos 1 y 2 estamos en presencia de una **investigación de desarrollo** donde se aplican teorías ya existentes a situaciones nuevas, pero que no requiere la elaboración de un nuevo modelo teórico. Existen otras investigaciones aplicadas en donde se aplican diversos métodos del campo científico como modelos empíricos o matemáticos a la esfera específica que se investiga con objetivos muy bien definidos

En los casos 3, 4 y 5 en donde las teorías existentes no explican a plenitud el problema investigado, donde se requiere acudir a elementos teóricos generales y a desarrollar una nueva teoría que de solución a dicho problema se dicen entonces que estamos en una **investigación científica** que necesita de la imaginación y creatividad del investigador.

La elaboración del marco teórico no es tan sólo reunir información, sino que la relaciona, integra y sistematiza, contribuyendo en alguna medida a la conformación de un nuevo modelo teórico, esto no es un aporte teórico pero si requiere de la capacidad de síntesis y enfoque del investigador.

Existen diferencias entre el Marco conceptual y el Marco Teórico, el **Marco Conceptual** se refiere a las características del medio, de aquello que precisa todo lo que rodea al objeto de investigación o su campo de acción, aquí el investigador define y delimita según su criterio y de acuerdo a su marco teórico **conceptos** involucrados en las variables de investigación, el **Marco Teórico** es lo existente sobre el objeto o campo, es la descripción de los elementos teóricos planteados por una y o por diferentes autores y que permiten al investigador fundamentar su proceso de conocimiento.

Se llama **Marco de referencias**, al resultado del planteamiento del problema y objetivos de la investigación, para formularlo hay que tener definido primero el **Marco Teórico**. Para hacer el **Marco Teórico** usted debe: revisar la bibliografía básica, seleccionar la bibliografía más importante, hacer las fichas o resúmenes de los autores más relevantes, etc. Para hacer el Marco Conceptual usted debe: seleccionar los términos a emplear en la investigación haciendo un glosario de dichos términos. Hay algunas corriente que hacen coincidir el Marco de Referencia con el Marco Teórico.

La elaboración del Modelo teórico adquiere importancia pues en él se establecen las conexiones con el problema, los objetivos y los métodos que se utilizan para llevar a cabo la investigación, se caracteriza mediante: los conceptos, las leyes, las teorías, modelos, cuadros.

1.4. Tipos de Estudios en la Investigación:

Cuando se decide que se va a realizar la investigación y bajo cualquier enfoque que utilicemos es muy importante visualizar el alcance del estudio que se va a llevar a efecto, según Danhke (1089) se dividen los tipos de estudios en **exploratorios, descriptivos, correlacionales y explicativos.**

Exploratorios

Los estudios **exploratorios** sirven para preparar el terreno y por lo común anteceden a los otros tres tipos. Los estudios descriptivos por lo general fundamentan las investigaciones correlacionales, las cuales a su vez proporcionan información para llevar a cabo estudios explicativos que generan un sentido de entendimiento y son altamente estructurados.

Los estudios exploratorios se efectúan, normalmente, cuando el objetivo es examinar un tema o problema de investigación poco estudiado, del cual se tienen muchas dudas o no se ha abordado antes. Si la revisión de la literatura reveló que tan sólo hay guías no investigadas e ideas vagamente relacionadas con el problema de estudio, o bien, si deseamos indagar sobre temas y áreas desde nuevas perspectivas o ampliar las existentes. Los estudios exploratorios sirven para familiarizarse con fenómenos relativamente desconocidos.

Descriptivos

Los estudios **descriptivos** buscan especificar las propiedades, las características y los perfiles importantes de personas, grupos, comunidades o cualquier otro fenómeno que se someta a un análisis. Describen situaciones, eventos o hechos, recolectando datos sobre una serie de cuestiones y se efectúan mediciones sobre ellas, buscan especificar propiedades, características y rasgos importantes de cualquier

fenómeno que se analice. Estos estudios presentan correlaciones muy incipientes o poco elaboradas.

Correlacionales

Los estudios **correlacionales** tienen como propósito evaluar la relación que existe entre dos o más conceptos, categorías o variables. En el caso de estudios correlacionales cuantitativos se mide el grado de relación entre dos o más variables que intervienen en el estudio para luego medir y analizar esas correlaciones y evaluar sus resultados.. La utilidad principal de los estudios correlacionales cuantitativos son saber cómo se puede comportar un concepto o una variable conociendo el comportamiento de otras variables relacionadas.

Explicativos

Los estudios **explicativos** van más allá de la descripción de conceptos o fenómenos o del establecimiento de relaciones entre conceptos, están dirigidos a responder a las causas de los eventos, sucesos y fenómenos físicos o sociales. Las investigaciones explicativas son más estructuradas que las demás clases de estudios e implican los propósitos de ellas.

Las investigaciones cualitativas se inician en su mayoría con enfoques exploratorios o descriptivos pero se plantean con alcances correlacionales, sin estadísticas, o explicativos. En las investigaciones cuantitativas se pueden aplicar cualesquiera de los estudios antes explicados, una investigación puede iniciarse inclusive como exploratoria y acabar siendo explicativa.

1.5. Las Hipótesis en la Investigación.

Para hacer el planteamiento correcto acerca de la solución de un problema científico es necesario la formulación de determinadas suposiciones o predicción, que tiene como punto de partida los conocimientos teóricos y empíricos existentes sobre los hechos y fenómenos que dan origen al problema planteado (marco teórico). Son proposiciones de carácter afirmativo, en su gran mayoría, que el

investigador plantea con el propósito de llegar a explicar hechos o fenómenos que caracterizan o identifican el objeto del conocimiento.

Una hipótesis es una suposición científicamente fundamentada y novedosa acerca de las relaciones y nexos existentes de los elementos que conforman el objeto de estudio y mediante la cual se le da solución al problema de investigación y que constituye lo esencial del modelo teórico concebido.

Las hipótesis no necesariamente son formuladas en todas las investigaciones, en las investigaciones de tipo cuantitativo, cuyo método es el deductivo siempre se formulan hipótesis si su carácter es correlacional, explicativo o cuando se hacen estudios descriptivos y se desean hacer pronósticos de cifras o hechos. En investigaciones cualitativas, por lo general no se formulan hipótesis antes de la recolección de los datos, en los casos más inductivos, pero cuando su alcance es correlacional o explicativo se pueden formular las hipótesis durante la obtención de la información, después de recabar los datos, al analizarlos o al establecer las conclusiones.

Forma lógica matemática de una Hipótesis:

En forma de lógica matemática las hipótesis se formulan de la siguiente forma: **Si p entonces q**, donde p y q son proposiciones que pueden ser afirmativas o negativas, pero que generalmente son positivas en el caso de las hipótesis, traducido a lenguaje normal sería: si hacemos esto, entonces sucederá esto otro. Las hipótesis son sólo proposiciones sujetas a comprobación empírica y a verificación en la realidad (para la investigación cuantitativa) y observación en el campo (para la investigación cualitativa)

La hipótesis junto con el problema y el objetivo cumple una labor de orientación fundamental, ya que la solución del problema y la demostración o no de la hipótesis de trabajo van a definir las tareas en todo este proceso, con vistas a lograr el objetivo, fundamentalmente en las investigaciones de tipo cuantitativo. La formulación de la hipótesis constituye un proceso del conocimiento hacia la posible ley desde el punto de vista formal es una conjetura o suposición que se expresa en forma de enunciado afirmativo y que generalmente enlaza al menos dos

elementos que denominamos características, variables, indicadores, propiedades, etc. La ausencia de la hipótesis empobrece el nivel teórico de la investigación y no va a permitir una orientación adecuada sobre el tipo de datos necesarios a buscar y reduce el modelo teórico del objeto investigado. En las hipótesis como predicción, suposición, proposición se dejan sentadas las posibles causas que generaron el problema: se establecen las variables, las relaciones entre ellas y se prevén los métodos a utilizar en la investigación. Esto hace de la **hipótesis el elemento rector del proceso de Investigación Científica**.

La formulación de las hipótesis tiene cuatro aspectos o fuentes a tener presentes:

1. *Aspecto metodológico*. Señala el nivel de profundidad con el cual el investigador busca abordar el objeto de conocimiento. De la teoría o sistema de conocimientos debidamente organizados y sistematizados, lo que se deriva mediante un proceso de deducción lógica.
2. *Estudio exploratorio*. Primer nivel de conocimiento, permite al investigador familiarizarse con el fenómeno que se investiga. Es el punto de partida para la formulación de otras investigaciones con el mayor nivel de profundidad. De la observación de los hechos o fenómenos concretos y sus posibles relaciones, mediante un proceso inductivo.
3. *Estudio descriptivo*. Según el nivel de conocimiento. Identifica características del universo de investigación, señala formas de conducta, establece comportamientos concretos y descubre y comprueba asociación entre variables. De la información empírica disponible la cual puede provenir de diferentes fuentes de otras investigaciones sobre el problema, de la experiencia que posee el propio investigador el cual tiene un papel relevante, en tanto que *es el que en un plano subjetivo, especulativo la propone*.
4. *Estudio explicativo*. Tercer nivel de conocimiento. Orienta la comprobación de hipótesis causales.

Las Variables.

Una variable es una propiedad que puede variar y cuya variación es susceptible de medirse u observarse.. Las hipótesis pueden ser más o menos generales o precisas, e involucran dos o más variables.

Hay muchas clasificaciones en cuanto a los tipos de hipótesis que existen, algunos las clasifican como:

1. *Hipótesis de trabajo*: las que se plantea el investigador en la fase exploratoria con insuficiente conocimiento del proceso.
2. *Hipótesis Real*: es el resultado de un conocimiento profundo de los vínculos y relaciones existentes entre los elementos presentes en el objeto de estudio, de la profundización de un marco teórico donde se desarrolla la investigación y del propio modelo teórico.
3. *Hipótesis descriptivas*: son las que reflejan el comportamiento de una variable relativamente superficial o fenoménica n el objeto de investigación.
4. *Hipótesis explicativas*: establecen una relación de dependencia causal entre una o varias de las variables.

Clasificación general de las Hipótesis.

Una clasificación general del enfoque cuantitativo fundamentalmente es la siguiente:

1. **Hipótesis de Investigación**. Son combinaciones de las hipótesis 1, 2 y 3 dadas anteriormente, proposiciones tentativas acerca de las posibles relaciones entre dos o más variables del problema de investigación. Pueden ser **descriptivas**, se describe alguna variable en términos de valores prefijados. Hipótesis **correlacionales** que especifican relaciones entre dos o más variables, no sólo establecen si dos o más variables están correlacionadas sino cómo lo están. Hipótesis de la **diferencia entre grupos**. Tienen el objetivo de hacer comparaciones entre grupos que intervienen en la investigación. **Hipótesis de relaciones de causalidad**. Establecen relaciones causa-efecto, no sólo afirman las relaciones entre las

variables y cómo se dan éstas, sino que además proponen un sentido de entendimiento entre ellas.

2. **Hipótesis Nulas.** Son en cierta forma el reverso de las hipótesis de investigación, constituyen proposiciones acerca de la relación entre variables sólo que sirven para refutar o negar lo que afirma la hipótesis de investigación. Se les denota como H_0 .
3. **Hipótesis Alternativas.** Reflejan lo contrario a las hipótesis nulas, posibilidades alternas ante las hipótesis nulas o de investigación. Se les denota como H_1 .
4. **Hipótesis Estadísticas.** Exclusivas del enfoque cuantitativo y representan la transformación de las hipótesis de investigación, nulas y alternativas en símbolos estadísticos. Estos símbolos pueden ser de \geq , de \leq (unilaterales) o de signos de $=$ (bilaterales).

1.6. Esquema Básico de un Proyecto de Investigación.

Titulo de la Investigación. (Titulo, nombre de autores, organización).

Resumen. (Breve reporte de lo que se quiere hacer)

Planteamiento del Problema.

Antecedentes del Tema. (Capacidad investigadora del grupo de trabajo)

Marco Teórico.

Hipótesis.

Metodología. (Diseño, población y muestra, técnicas, índice analítico, guía de trabajo)

Plan de análisis de los resultados esperados.

Referencias bibliográficas.

Cronograma. (Plan de actividades, diagrama de Gantt)

Presupuesto. (Costos, recursos, finanzas)

Anexos.

Esquema Básico de un Informe de Investigación Terminada.

Titulo de la Investigación. (Titulo, nombre de autores, organización y fecha)

Resumen. (Breve reporte del contenido de la investigación)

Introducción. (Antecedentes, problema de investigación, limitaciones, resultados)

Planteamiento del Problema.

Marco teórico.

Hipótesis.

Metodología. (Diseño, población y muestra, técnicas, índice analítico, guía de trabajo)

Plan de análisis de los resultados.

Conclusiones y recomendaciones.

Referencias bibliográficas.

Anexos.

1.7.- Los Diseños de la Investigación:

La manera de enfrentar la investigación

Seleccionar un diseño de investigación significa trazar un plan estratégico para responder a las preguntas de investigación planteadas al inicio.

La literatura especializada sobre metodología de la investigación plantea diferentes clasificaciones acerca de los tipos de diseños existentes. Debemos aclarar que en general, no se considera que ningún tipo de investigación sea superior que el otro, en muchas ocasiones en una misma investigación se aplican diferentes diseños.

Diseños no experimentales de investigación.

La investigación no experimental es la que no manipula deliberadamente las variables a estudiar. Lo que hace este tipo de investigación es observar fenómenos tal y como se dan en su contexto actual, para después analizarlo.

En un estudio no experimental no se construye ninguna situación, sino que se observan situaciones ya existentes.

Existen diferentes criterios para clasificar la investigación no experimental, adoptaremos la dimensión temporal, es decir de acuerdo con el número de momentos o puntos en el tiempo en los cuales se recolectan los datos.

Criterios de clasificación.

Según este criterio las investigaciones no experimentales pueden ser: **longitudinales (datos de panel) y transversales.**

Investigación transversal.

Recolectan los datos en un solo momento, en un tiempo único. Su propósito es describir variables y analizar su incidencia e interrelación en un momento dado.

Este tipo de investigaciones es como una fotografía en un momento dado del problema que se está estudiando y puede ser: descriptiva o de correlación, según el problema en estudio.

Investigación longitudinal.

A diferencia de la investigación transversal en este tipo de estudios se obtienen datos de la misma población en diferentes momentos. Aquí se comparan los datos obtenidos en las diferentes oportunidades a la misma población o muestra y se pueden analizar los cambios a través del tiempo de determinadas variables o en las relaciones entre ellas.

Este tipo de estudios se utiliza mucho en el área de investigación de mercados para el análisis en los cambios de hábitos de compra y

comportamiento de los consumidores. También en el área de la Psicología, para estudiar el cambio en los pacientes después de aplicar tal o cual tratamiento.

Ejemplo: Se quiere analizar la productividad de los trabajadores en una empresa X y para ello se toman datos durante dos años trimestralmente.

Comparación de los diseños longitudinales y transversales.

Los estudios longitudinales tienen la ventaja con respecto a los transversales, que nos proporcionan información sobre la manera de cómo evolucionan las variables y las relaciones entre ellas en el tiempo, pero esto implica mayores costos.

1.7.1.- Diseños experimentales.

Los investigadores realizan experimentos virtualmente en todos los campos del saber humano, por lo general para describir algo acerca de un proceso o un sistema.

En los experimentos se diseñan pruebas en las cuales se inducen cambios es decir se manipulan las variables que intervienen en un proceso o sistema deliberadamente (supuestas causas), de manera que sea posible observar, identificar y analizar las causas en la respuesta obtenida.

Por ejemplo: En las actividades docentes se realizan investigaciones con varios grupos de estudiantes para determinar una nueva experiencia pedagógica. En este caso al grupo que se expone a la presencia de la variable independiente se le conoce como grupo experimental y al grupo en el cual está ausente se le denomina grupo de control, aunque ambos grupos participen en el experimento.

¿Cuáles son las características de la investigación no experimental en comparación con la investigación experimental?

Ambas formas de investigación son muy valiosas para el desarrollo del conocimiento humano, y ningún tipo es mejor que el otro. El diseño a seleccionar dependerá del tipo de problema que se desea resolver y el contexto de estudio.

En los estudios experimentales el control sobre las variables es mucho más riguroso que en los no experimentales, donde las que intervienen no son para nada manipuladas por el investigador.

En los estudios no experimentales, resulta muy complejo separar los efectos de las múltiples variables que intervienen en el estudio y saber cuánto contribuyó cada una, situación que no sucede en los estudios experimentales donde podemos estudiar las variables por separado o conjuntamente con otras para conocer sus efectos.

En los estudios experimentales las variables independientes tienen menor fuerza que en la realidad. Es decir, en los laboratorios dichas variables no muestran la magnitud real de sus efectos. En la investigación no experimental las variables tomadas son más reales y consecuentemente tendremos mayor validez externa.

¿Qué relación existe entre el tipo de estudio, las hipótesis y el diseño de la investigación?

El planteamiento del problema, y el marco teórico nos indican si un estudio se iniciará con fines exploratorios, descriptivos, correlacionales o explicativos. Asimismo, el tipo de estudio nos indicará la forma de plantearnos las hipótesis, y sobre la base de ello se seleccionará el tipo de diseño de investigación a aplicar.

Es decir el tipo de diseño a elegir se encuentra condicionado por el problema a investigar, el contexto que rodea la investigación, el tipo de estudio a efectuar y la forma de haber planteado las hipótesis.

1.8.- Las Variables en las Investigaciones:

Las **Variables** ya han sido estudiadas anteriormente cuando se vieron las hipótesis, se denomina **variable** a cualquier característica que pueda ser percibida (o medida) y que cambie de un sujeto a otro, o en el mismo sujeto a lo largo del tiempo.

Por lo tanto, puede definirse **variable** de diferentes maneras:

- Todo aquello que puede ser medido, observado y manipulado durante un estudio.
- Cualquier característica que varía de un miembro a otro en una población determinada.
- Cualquier cualidad o característica, constituyente de una persona o cosa, que es susceptible de ser medida y que está sujeta a cambio.

Tipos de Variables.

Las variables pueden ser clasificadas de diversos modos:

- Según la naturaleza de la característica medida,
- Según la manera de medir dicha característica,
- Según la escala de medición empleada, y
- Según su relación con otras variables.

Según su Naturaleza

De acuerdo con la característica medida, una variable puede ser:

- **Discreta:** si los valores u observaciones que pertenecen a ella son distintas y separadas, es decir que pueden ser contadas. Ejemplos: el número de años que vive una persona
- **Continua:** es una característica que puede adoptar infinito número de valores a lo largo de un continuo. Magnitudes físicas como la masa, la velocidad, la temperatura, etc., son ejemplos de variables continuas. El peso de una persona puede variar a lo largo de un número infinito de valores posibles. La cantidad de valores intermedios entre dos puntos de la medición, por ejemplo: 50 y 51 Kilos, es ilimitada y depende de la precisión con que se pueda efectuar la medición.

Según la Manera de ser Medidas:

Las variables pueden ser medidas de forma **cualitativa**, cuando la característica es descrita en términos de una cualidad específica, sin asociarle valores numéricos.

Por el contrario, las variables pueden ser **cuantificadas (cuantitativas)** en términos numéricos.

Las variables de naturaleza discreta suelen medirse en términos cualitativos. Hay, sin embargo, dos circunstancias en las cuales se utilizan modos cuantitativos: la primera es cuando el fenómeno medido no es una característica de un solo elemento (por ejemplo, el color), sino la frecuencia de ocurrencia de un evento. Estas variables son las llamadas **numéricas discretas o variables de conteo**.

Según la Escala de Medición:

A las variables medidas de forma cualitativa se les pueden aplicar diferentes escalas de medición. Esta escala de medición (cualitativa) de características discretas se llama **nominal**.

Un conjunto de datos se llama **nominal** si a los valores u observaciones que pertenecen a él se les puede asignar un código, en la forma de un número, donde los números son simplemente una etiqueta. Los datos nominales pueden ser contados, pero no pueden ser ordenados o medidos. Ejemplo: los varones pueden ser codificados como “0” y las mujeres como “1”.

Un conjunto de datos se denomina **ordinal** si a los valores u observaciones que pertenecen a él se les puede asignar un orden o asociar una escala. Los datos ordinales pueden ser contados y ordenados, pero no pueden ser medidos.

Las categorías, para un conjunto ordinal, deben tener un orden natural; por ejemplo, aspectos en la escala de excelente, buena, regular, mala y pésima. Un puntaje de 5 indica mayor calidad que un puntaje de 4. Así que los datos resultantes son ordinales.

Según su Relación con otras Variables:

En el lenguaje de los diseños experimentales, al factor que se espera que sea la causa de algo y que es manipulado por los investigadores se le denomina **variable independiente (VI)**, justamente porque sus valores no dependen de otras variables sino de otros aspectos a los cuales no se les puede influir.

La conducta que se verá influida por la variable independiente se llama **variable dependiente (VD)**, porque sus valores “dependerán” de aquellos que hubiésemos elegido para la variable independiente.

Las parejas de variables independientes-dependientes no son siempre monógamas. Una variable independiente puede emparejarse con más de una dependiente en cuyo caso estaremos en procesos multivariados.

1.9.- Un Enfoque a la Investigación Cualitativa.

La investigación cualitativa tiene significados diferentes en cada momento. Una primera definición, aportada por Denzin y Lincoln (1994:2) destaca que:

“Es multimetódica en el enfoque interpretativo, naturalista hacia su objeto de estudio”.

Esto significa que los investigadores cualitativos estudian la realidad en su **contexto natural**, tal y como sucede intentando sacar sentido de..., o interpretar los fenómenos de acuerdo con los significados para las personas implicadas.

Taylor y Bogdan (1986:20) consideran, en un sentido amplio, la investigación cualitativa como:

“Aquella que produce datos descriptivos: las propias palabras de las personas, habladas o escritas, y la conducta observable”.

Por su parte Miles y Huberman (1994:5-8), consideran como características básicas de la investigación naturalista las siguientes:

- 1. “Se realiza a través de un prolongado e intenso contacto con el campo o situación de vida.*
- 2. El papel del investigador es alcanzar una visión holística (sistemática, amplia, integrada) del contexto objeto de estudio.*
- 3. El investigador intenta capturar los datos sobre las percepciones de los actores desde dentro, a través de un proceso de profunda*

- atención, de comprensión empática y de suspensión o ruptura de las preconcepciones sobre los tópicos objeto de discusión.*
4. *Leyendo a través de estos materiales, el investigador puede aislar ciertos temas y expresiones que pueden revisarse con los informantes, pero que deberían mantenerse en su formato original a través del estudio.*
 5. *Una tarea fundamental es la de explicar las formas en que las personas en situaciones particulares comprenden, narran, actúan, y manejan sus situaciones cotidianas.*
 6. *Son posibles muchas interpretaciones de estos materiales, pero algunas son convincentes por razones teóricas o consistencia interna.*
 7. *Se utilizan relativamente pocos instrumentos estandarizados. El investigador es el principal instrumento de medida.*
 8. *La mayor parte de los análisis se realizan con palabras”.*

En tal sentido LeCompte (1995:67) considera que la misma podría entenderse como:

“Una categoría de diseños de investigación que extraen descripciones a partir de observaciones que adoptan la forma de entrevistas, narraciones, notas de campo, grabaciones, descripciones, registros escritos de todo tipo, fotografía o películas o artefactos”.

Por su parte Stake (1995:87) sitúa las diferencias fundamentales entre la investigación cualitativa y cuantitativa en tres aspectos básicos:

1. *La distinción entre la explicación y la comprensión como propósito del proceso de indagación; el objetivo de la investigación cualitativa es la comprensión, centrandó la indagación en los hechos; mientras que la investigación cuantitativa fundamentará su búsqueda en las causas, persiguiendo el control y la explicación.*
2. *La distinción entre el papel personal e impersonal que puede adoptar el investigador; se valora el papel personal que asume el*

investigador desde el comienzo de la investigación, interpretando los sucesos y acontecimientos, frente a la posición mantenida desde los diseños cuantitativos, donde el investigador interpreta los datos una vez que se han recogido y analizado estadísticamente. En la investigación cualitativa se espera una **descripción densa, una comprensión experiencial y múltiples realidades.**

3. *La distinción entre conocimiento descubierto y conocimiento construido*, en la investigación cualitativa, el investigador no descubre, sino construye el conocimiento.

Como síntesis de su perspectiva, Stake considera como aspectos diferenciales de un estudio cualitativo su carácter holístico, empírico, interpretativo y empático.

Por todo lo anteriormente planteado podemos señalar que la investigación cualitativa al permitir una descripción exhaustiva y densa de la realidad concreta objeto de investigación puede ser utilizada no sólo en educación, sino en psicología, economía, medicina, antropología, etc.

Finalmente conviene señalar que el enfoque cualitativo pretende ofrecer profundidad detallada del objeto de estudio mediante una descripción densa y registro cuidadoso de los datos, con el fin de obtener una coherencia lógica durante el suceso de los hechos.

1.9.1.- Técnicas para la Recogida de Datos.

La Observación Como Método de Recogida de Datos.

Entre los diferentes métodos que nos sirven para recoger información de la vida social tenemos la observación.

La observación es una de las técnicas cualitativas más aplicada en la etnografía y precisamente en el marco educativo, por la riqueza de su información y la influencia de la misma en la formación del estudiante durante el proceso de enseñanza-aprendizaje.

Contemplar sistemática y detenidamente cómo se desarrolla la vida social, sin manipularla ni modificarla, tal cual ella discurre por sí misma, es el método de la observación (Olabuénaga e Ispízu, 1989:80).

La observación, por principio, es susceptible de ser aplicada a cualquier conducta o situación. Pero una observación sin una adecuada planificación pierde interés y los resultados no son utilizados con efectividad. Como plantea Whitehead (1967:28): "*Saber observar es saber seleccionar*", es decir plantearse previamente qué es lo que interesa observar, plantearse una estructura teórica previa o esquema conceptual.

Observamos las conductas y las conversaciones, la participación y el retraimiento, la comunicación y el silencio de las personas. Esta observación común y generalizada puede transformarse en una poderosa herramienta de investigación social y en técnica científica de recogida de información si se efectúa (Olabuénaga e Ispízu, 1989:79-80):

- Orientándola y enfocándola a un objetivo concreto de investigación, formulado de antemano.
- Planificándola sistemáticamente en fases, aspectos, lugares y personas.
- Controlándola y relacionándola con proposiciones y teorías sociales, planteamientos científicos y explicaciones profundas.
- Sometiéndola a controles de veracidad, de objetividad, de fiabilidad y de precisión.

Tipos de Observación Cualitativa

En el marco de la investigación cualitativa podemos encontrar cuatro tipos de observación que nos posibilitan la recogida de datos y que, según Olabuénaga e Ispízu (1989:83), son:

1. Observación panorámica-participante.
2. Observación panorámica-no participante.
3. Observación selectiva-participante.
4. Observación no selectiva-participante.

Esta clasificación ha tenido en cuenta el grado de participación del observador, el cual podrá elegir o combinar según su objeto de estudio, pero generalmente muchos autores entre los que podemos citar: Walker (1989:143) y Pérez Serrano (1994:23) entre otros coinciden en agruparla en dos grandes grupos: participante y no participante.

En los primeros días los investigadores-observadores deben permanecer comportándose como miembros activos del colectivo y dedicar el mayor tiempo posible que le permita adentrarse en el contexto, para sentir lo que es vivir dicha situación, lo que posibilita una total entrega a la tarea de interpretación, donde se identifica y, al mismo tiempo permanece distante del contexto observado. Este tipo de investigación sirve para obtener de los individuos sus definiciones de la realidad y los constructos que organizan su mundo (Goetz y LeCompte, 1988:126).

Taylor y Bogdan (1992:75) plantean:

"Las notas de campo deben incluir descripciones de personas, acontecimientos y conversaciones, tanto como las acciones, sentimientos, intuiciones o hipótesis de trabajo del observador. La secuencia y duración de los acontecimientos y conversaciones se registra con la mayor precisión posible. La estructura del escenario se describe detalladamente. En resumen, las notas de campo procuran registrar en el papel todo lo que se puede recordar sobre la observación. Una buena regla establece que si no está escrito, no sucedió nunca".

La Encuesta.

Los etnógrafos utilizan la encuesta como técnica que le posibilita descubrir los componentes de los mundos de sus participantes y los constructos con arreglo a los cuales esos mundos están estructurados.

Discusión Grupal.

La discusión de grupo es una técnica de investigación que consiste en reunir a un grupo de seis a diez personas y suscitar entre ellas una conversación sobre el tema que queremos investigar, la cual debe estar

dirigida en nuestro caso por uno de los integrantes del equipo de trabajo, con vistas a tomar notas y no dejar escapar ningún detalle útil para el desarrollo del mismo. En su obra, “Tratado de Sociología empírica”, Wener Mangold (1973:249) muy oportunamente señala que el empuje efectivo para la aplicación de las discusiones de grupo, en lugar de entrevistas aisladas, se origina en la reflexión de que opiniones y actitudes sólo raras veces se transforman en la realidad en un comportamiento social relevante bajo condiciones que corresponden al modelo de la escuela. Esta técnica, que se aplica muy frecuentemente a estudios de sociología, busca descubrir cómo influyen las opiniones en las conductas de las personas, tanto para determinarlas como para modificarlas.

La Entrevista.

La entrevista es un instrumento fundamental en las investigaciones sociales, pues a través de ella se puede recoger información de muy diversos ámbitos relacionados con un problema que se investiga, la persona entrevistada, su familia, y el ambiente en que se halla inmersa.

Consideramos muy importantes los aspectos que señala Woods (1987: 77), cuando se refiere a que las entrevistas etnográficas por sí mismas tienen un carácter muy especial, algo afín a la observación participante. Además resalta que el éxito o fracaso de esta técnica depende directamente de la persona y de la disposición del entrevistador. En este sentido es significativo señalar que los atributos personales que exige la entrevista son los mismos que en otros aspectos de la investigación y giran siempre en torno a la confianza, la curiosidad, la naturalidad, es decir promover en este ámbito una adecuada interacción, que favorezca un vínculo de amistad, un sentimiento de solidaridad y unión en busca de la solución de la problemática objeto de estudio.

El éxito de esta técnica radica en la comunicación personal, en el vínculo que se establezca persona a persona. Algunos autores como Goetz y LeCompte (1988), Pérez Serrano (1994), Walker (1989), Álvarez Rojo

(1989), Olabuénaga e Ispízu (1989), consideran algunos elementos a tener en cuenta para planificar una entrevista, los que hemos analizado y adaptado, resumiéndolos de la siguiente forma:

- Adecuada selección del contenido y estructuración adecuada de las preguntas, insistir en las preguntas abiertas.
- Crear un clima favorable y hacer que la gente se sienta cómoda.
- Realizar una adecuada orientación, tanto en el marco de la entrevista como en el contenido de la misma.
- Propiciar una adecuada comunicación, siempre tener planificadas las preguntas, con cierta flexibilidad, nunca caer en la improvisación.
- La conversación no sigue un esquema rígido de desarrollo, razón por la cual, es posible retroceder, retomar temas ya tratados.
- Mantener la conversación con incentivos que garanticen e incrementen la motivación, el interés y la participación espontánea.
- La amistad no debe interferir en el carácter profesional de la entrevista.
- Relación amistosa entre entrevistador y entrevistado.
- En algunos momentos se hace necesario cambiar de actividad, para luego retomar la misma.
- Concretar progresivamente la conversación para obtener datos cada vez más reveladores, más significativos.

Durante el desarrollo de la entrevista el entrevistador juega un papel fundamental, pues debe intervenir de forma tal que oriente o influya en la respuesta del entrevistado. Se trata de una forma de sugestión que se da la mayoría de las veces de forma inconsciente, tanto por parte del entrevistador como por parte del entrevistado.

Fuentes de Rigor en la Investigación Cualitativa

Cuando hablamos de calidad de la investigación aludimos al rigor metodológico con que ha sido diseñada y desarrollada, y a la confianza que, como consecuencia de ello, podemos tener en la veracidad de los resultados conseguidos. En general, la idea de calidad de la investigación se asocia por tanto a la credibilidad del trabajo desarrollado.

A continuación exponemos los términos apropiados para definir el rigor científico, contraponiendo la investigación de tipo cuantitativo y cualitativo.

ASPECTO	TERMINO CIENTIFICO	TÉRMINO NATURALÍSTICO.
Valor verdadero.	Validez interna.	Credibilidad.
Aplicabilidad.	Validez externa (generalización).	Transferencia (transferir los resultados a otros contextos).
Consistencia.	Fiabilidad.	Dependencia.
Neutralidad.	Objetividad.	Confirmación

Resumen de los aspectos a considerar para fundamentar el rigor científico de la investigación cualitativa (Guba, 1981:104).

Hay que tener muy en cuenta como todos estos aspectos pueden mezclarse e interrelacionarse en la medida que se valoren íntegramente todas las técnicas que se hayan utilizado, pero sí está muy claro que las vías que propician el valor científico en la investigación cualitativa se resumen en:

- Credibilidad o valor de verdad.
- Transferencia o aplicabilidad.
- Consistencia o Dependencia.
- Confirmación.

En busca del rigor científico de todos los datos obtenidos, se consideran las aportaciones de Goetz y LeCompte, (1988:215) sintetizadas en el siguiente cuadro:

BÚSQUEDA DEL RIGOR CIENTIFICO.

Credibilidad (validez interna):

- Aumentar la probabilidad de que los datos hallados sean creíbles. Observación persistente, diarios de experiencias, encuestas, análisis de documentos, discusión grupal y triangulación.
- *Contrastar los resultados con las "fuentes"*.
- Negociación inicial y durante todo el proceso.

Transferencia (validez externa):

- Actuar por parecidos contextuales y descripción densa.

Consistencia (replicabilidad):

- Triangulación de investigadores, de métodos y de resultados.
- Establecer pistas de revisión a través de los diarios de experiencias, informes de investigación, análisis de documentos etc.

Confirmabilidad (fiabilidad externa):

- Son considerados válidos diferentes métodos como: Observación, diario, encuestas, análisis de documentos, discusión grupal, triangulación, etc.

Resumen de los principales aspectos que apoyan el rigor científico en el presente trabajo.

Métodos Utilizados para Corroborar el Rigor en la Investigación Cualitativa.

Los métodos para corroborar la fiabilidad de sus resultados son:

Trabajo prolongado en el mismo lugar: Se hace necesario disponer del tiempo suficiente que posibilite al investigador integrarse al escenario de actuación para de esta forma evitar, tanto como sea posible, las distorsiones producidas por la novedad de su presencia, el temor o las expectativas que ha podido ocasionar en la etapa inicial, (son aspectos a considerar en los primeros momentos) por lo que es necesario dedicar un extenso período a un mismo lugar para que los investigados se acostumbren a la presencia de investigadores y se convenzan de que no constituyen una amenaza (es muy posible que ese factor se atenúe o desaparezca con el tiempo). Finalmente queremos señalar que la observación participante tiene que ser prolongada para poder separar lo anecdótico de lo relevante, lo irrelevante de lo sustantivo, lo superficial de lo profundo (Santos, 1990:165).

Separación periódica de la situación: La separación temporal del investigador durante el desarrollo de la investigación le permite tomar perspectiva y evita el peligro de su conversión en nativo (Santos, 1990:165). Es oportuno ocupar estos espacios temporales, en la redacción de informes, análisis de los resultados u otras actividades que en cierto sentido condicione la asimilación integrada del objeto de investigación y poder contrastar las mismas en condiciones similares. La duración de estas etapas de ausencias no debe ser larga, para evitar que el investigador pierda la secuencia de trabajo y la vinculación a la experiencia.

Triangulación de Investigadores: La realidad del objeto de investigación puede ser observada por todos los que se impliquen de una forma u otra en la misma, posibilitando la confrontación de diferentes percepciones con el fin de reflexionar e interpretar los datos obtenidos. No se puede pensar que todos los investigadores tengan una misma perspectiva de análisis, sencillamente es interesante confrontar varias opiniones para enriquecer el análisis de los resultados. Guba (1981:148) plantea:

"No se debería aceptar ningún ítem de información que no pueda ser verificado por dos fuentes al menos".

Es importante en el marco de esta etapa no utilizar etiquetas clasificadoras, ni siquiera las de carácter sociológico (de pertenencia). Nos interesa, sencillamente, contrastar las opiniones (Santos, 1990:166).

Triangulación de Métodos: Goetz y LeCompte (1988:214) señalan que la mejor defensa contra las amenazas a la fiabilidad interna de los estudios etnográficos es la presencia de más de un investigador en el campo, pues esto nos posibilita enriquecer el análisis a realizar, garantizando una mayor fiabilidad de los resultados. y en tal sentido se debe contrastar sistemáticamente los resultados de los diferentes métodos.

Triangulación de los resultados: Su objetivo es provocar el intercambio de pareceres, comparar las diferentes perspectivas de los investigadores con los que se interpretan los acontecimientos que se han desarrollado en el lugar de trabajo. La confrontación de los resultados concede un valor de primer orden en el análisis reflexivo colectivo, analizando dichos resultados al hilo de la reflexión sobre la acción. El análisis de los resultados en una investigación cualitativa debe producirse en procesos de colaboración y contraste y diseminarse también mediante procesos de intercambio y participación en situaciones de igualdad de oportunidades.

Por su parte, Pérez Serrano (1994:90) realiza un resumen sobre los criterios y procedimientos que fundamentan el rigor científico en los datos obtenidos a través de la investigación cualitativa, los cuales ofrecemos a continuación en el siguiente cuadro:

CRITERIOS	PROCEDIMIENTOS
a) Valor de verdad: Isomorfismo entre los datos recogidos por el investigador y la realidad. Credibilidad.	<ul style="list-style-type: none"> ➤ Observación persistente. ➤ Encuesta. ➤ Triangulación. ➤ Comprobaciones con los participantes.
b) Aplicabilidad: Grado en que puede aplicarse los descubrimientos de una investigación a otros sujetos o contextos. Transferencia.	<ul style="list-style-type: none"> ➤ Muestreo teórico. ➤ Descripción exhaustiva. ➤ Recogida de abundantes datos.

CRITERIOS	PROCEDIMIENTOS
<p>c) Consistencia: Repetición de resultados cuando se realizan investigaciones en los mismos sujetos e igual contexto. Dependencia.</p>	<ul style="list-style-type: none"> ➤ Descripciones minuciosas de los informantes. ➤ Identificación y descripción de técnicas de análisis de datos. ➤ Delimitación del contexto físico y social. ➤ Pistas de revisión. ➤ Métodos solapados.
<p>d) Neutralidad: Garantía de que los descubrimientos de una investigación no están sesgados por motivaciones, intereses y perspectivas del investigador. Confirmación.</p>	<ul style="list-style-type: none"> ➤ Descripciones de baja inferencia. ➤ Comprobaciones de los participantes. ➤ Recogida mecánica de datos. ➤ Triangulación. ➤ Explicar posicionamiento del investigador.

Credibilidad.

Alude a la confianza en la veracidad de los descubrimientos realizados en una investigación y hace referencia a la necesidad de que exista un isomorfismo entre los resultados de la investigación y las percepciones que los sujetos participantes poseen de la realidad estudiada (Rodríguez Gómez, 1996:286).

La pieza clave en la calidad de la investigación se encuentra en el modo en que es narrada y en las evidencias que se presentan para apoyar su autenticidad. A través del informe de investigación se debe demostrar la veracidad de los argumentos, demostrando que existen datos que fundamentan dichos criterios.

La credibilidad hace referencia a que todos los datos de la investigación sean aceptables, es decir creíbles

Transferencia.

Consiste en transferir los resultados de la investigación a otros contextos. Por lo que se debe ofrecer una descripción densa que permita al lector manejar la información necesaria y suficiente que facilite establecer comparaciones y transferir dichos resultados.

Aunque la generalización no es posible dado al carácter único e irrepetible de los contextos, conductas y fenómenos que se estudian, sí que cabe la posibilidad de algún tipo de transferencia entre contextos, dependiendo del grado de similitud entre los mismos.

Dependencia o Consistencia

Los criterios de dependencia se garantizan a través de la triangulación de los resultados, pues se contrastan las perspectivas individuales de los investigadores y de los investigados con relación a la adquisición de nuevos conocimientos a través del trabajo diferenciado. También se contrastan los resultados obtenidos por los diferentes métodos aplicados, entre los que se pueden citar: la observación, el diario, encuestas abiertas, entrevista, las discusiones grupales y análisis de documentos, entre otras.

Confirmabilidad

Al igual que la dependencia, en la medida que se expone de forma clara y precisa los resultados del trabajo así como las vías utilizadas para lograrlo, se deja garantizada la fiabilidad externa o confirmabilidad. De esta manera a través de un análisis de los datos aportados en el cuerpo teórico del trabajo y en los anexos del mismo, dichos resultados pueden ser trasladados y ajustados a otros contextos concretos. En tal sentido la negociación sistemática entre todos los investigadores va generando fiabilidad, dejando todo clarificado en cada caso oportuno.

De igual manera, según Rodríguez Gómez (1996:288), la forma en que se intenta ofrecer evidencia de la confirmabilidad de los datos se basa, en el empleo de estrategias de triangulación, y sobre todo en el ejercicio de la **reflexión**.

En la medida que avance la investigación, la comprensión de la realidad objeto de estudio permite develar indicios y nuevos caminos a recorrer más, que presentar absoluta certeza.

Al respecto Santos (1990:169), citando a Guba (1981), señala que:

"Es dudoso que alguna vez surjan los criterios perfectos, hasta entonces, será prudente afirmar con humildad que se ha encontrado una nueva y más verdadera senda hacia el conocimiento".

El uso del paradigma naturalista o etnográfico está cargado de riesgos especiales para el investigador a causa de su carácter abierto, sin diseño inicial, no se puede anticipar los resultados, pero se considera que es preferible correr dicho riesgo, ya que al final del trabajo queda enriquecido en el sentido más amplio de la palabra, pues constituye una experiencia indescriptible para los que se inician en esta estrategia.

Capítulo 2: Sistemas de expertos no estructurados.

En la Etapa Inicial el o los expertos pueden utilizar una serie de métodos para organizar, agrupar, conformar y llegar a conformar consensos de opiniones. Son los llamados Métodos de grupos o métodos creativos de grupo, se utilizan métodos del tipo heurísticos donde juega un papel importante el uso de los llamados Métodos de Expertos. Entre los principales métodos creativos de grupos se encuentran:

- **Brain Storming.**
- **Grupos Nominales.**
- **Método Delphi.**

Se utilizan las técnicas del Pronóstico y el Método de Enfoque Genético que plantean deducir el comportamiento futuro partiendo del pasado. Pasaremos a describir estos métodos en forma general.

2.1.- El Método **Brain Storming** dirigido a la obtención de una gran cantidad de ideas del grupo de expertos reunidos, presenta las siguientes características:

- El grupo es dirigido por alguien con experiencia, llamado **facilitador**, que lo **excita**, mediante un plan, a la generación de ideas.
- En ocasiones se nombra a otra persona que registre las ideas obtenidas del experimento en un pizarrón, a esta persona se le llama **registrador**.
- Se aspira a un máximo de ideas para aumentar la probabilidad de respuestas útiles, que después pueden ser combinadas y perfeccionadas.
- Es ideal en grupo de expertos de pequeña cantidad.
- Se aplica a cuestiones bien definidas.
- La generación de ideas no es restringida.
- No se permite la crítica de una idea.
- El que dirige no da criterios.
- El análisis de las ideas se separa del que las generó.

Existen varias formas de llevarlo a cabo:

1. Las ideas se registran en una pizarra y finalmente, cuando no existan más ideas se agrupan detectando las que son repetitivas y se organizan como quedarían finalmente. El orden de presentación de las ideas es inmaterial.
2. Metaplan. Las ideas se anotan en tarjetas, de forma que todos las puedan, cuando un participante incorpora una tarjeta, lee las que ya hay colocadas, que a su vez le servirán de inspiración para nuevas ideas. Al final cada participante colorea en las tarjetas, con ideas de su interés, y finalmente las ideas más aceptadas son las que se escogen.
3. Seis-Tres-Cinco. Grupo de 6 participantes, cada uno escribe 3 propuestas en un papel y pasa la hoja al de al lado que inspirándose en esas ideas o soluciones, escribirá otras 3. Seis participantes, preparando tres ideas y analizando las de los cinco restantes dan 36 propuestas con un total de 108 sugerencias. (6 participantes * 3 ideas iniciales c/u = 18 ideas iniciales) * 5 rondas de los restantes participantes (18 * 5) = 90 propuestas + 18 iniciales = 108 sugerencias

2.2.- Grupos Nominales. Es una variante entre el Brain Storming y el Delphi, algunas características del mismo son:

1. Ideal para grupos de entre 10 a 20 expertos.
2. Se define abiertamente el objetivo.
3. Cada experto brinda sus ideas en forma anónima escrita.
4. No se debe evaluar ninguna idea hasta que todos los resultados se conozcan.
5. Deben debatirse todos los puntos antes de efectuar la votación.
6. Debe limitarse la discusión a la presentación de los pro y los contras de cada punto.
7. Puede permitirse la incorporación de una nueva idea a partir del análisis.
8. Utiliza siempre votación anónima.

9. Se realizan tantas votaciones como sean necesarias, para llegar a las ideas finalmente aprobadas por la mayoría.

2.3.- Método Delphi (Delfos oráculo de la Grecia Antigua), este método es la utilización sistemática del juicio intuitivo de un grupo de expertos para obtener un consenso de opinión. Sus principales características son:

- Anonimato.
- Retroalimentación controlada por el facilitador.
- Respuesta estadística de grupo. La información obtenida se procesa por medio de técnicas estadístico – matemáticas del diseño experimental.

Las Fases de su Aplicación son:

- Seleccionar el coordinador.
- Elaborar lista de candidatos a expertos que cumplan los requisitos predeterminados de experiencia, años de servicio, conocimientos sobre el tema, etc.
- Determinación del coeficiente de competencia de cada experto:

K comp. = $\frac{1}{2} (Kc + Ka)$ Coeficiente de competencia de cada experto

Donde: **Kc:** Coeficiente de Conocimiento. Se hace una encuesta en donde el candidato le otorga a cada una de las preguntas un valor, según el conocimiento que considere tenga al respecto. El coeficiente resulta del promedio de los valores que se otorga al candidato.

Ka: Coeficiente de Argumentación. Es la suma de los valores del grado de influencia de cada una de las fuentes de argumentación con respecto a una tabla patrón.

Ejemplo de una tabla patrón a manejar por el coordinador, en donde el experto debe poner el grado de influencia de cada una de las fuentes en su conocimiento y criterio según sean alto, medio o bajo:

Fuentes de Argumentación	Alto	Medio	Bajo
Análisis Teóricos realizados por usted	0.3	0.2	0.1
Experiencia obtenida	0.5	0.4	0.2
Trabajos de autores nacionales que conoce	0.05	0.04	0.03
Trabajos de autores extranjeros que conoce	0.05	0.04	0.03
Conocimientos propios sobre el estado del tema	0.05	0.04	0.03
Su intuición	0.05	0.04	0.03

Se suman todos los valores obtenidos y ese resultado forma el coeficiente de argumentación **Ka** de cada experto.

Dados los coeficientes **Kc** y **Ka** se calcula para cada experto el valor del coeficiente de competencia **K comp** siguiendo los criterios siguientes:

La competencia del experto es ALTA si **K comp > 0.8**
 La competencia del experto es MEDIA si **0.5 < K comp <= 0.8**
 La competencia del experto es BAJA si **K comp <= 0.5**

Se eligen los Expertos de entre los auto evaluados de Alta Competencia. Los expertos seleccionados no deben conocer a los restantes que fueron escogidos, todo debe ser hecho en forma individual, el método mantiene el anonimato y ese es uno de sus éxitos principales.

Las Rondas del Delphi:

- **Primera Ronda.** El facilitador debe escribir una pequeña explicación a los expertos por qué fue seleccionado y cuál es el objetivo del trabajo investigativo que se va a llevar a cabo, del cuál él forma parte por sus conocimientos y experiencia al respecto, a continuación se formula en forma clara y precisa la pregunta abierta para la que se quiere que él de sus criterios al respecto en forma escrita y lo responda por la vía de la correspondencia, fundamentalmente por e-mails en los tiempos actuales. Una vez recibidas las respuestas por parte del facilitador. Este puede construir una tabla en dónde las filas se reserven a los expertos supongamos m y las columnas a los criterios emitidos supongamos n. A esa tabla se le pondrá una X a la intersección del

experto con el criterio que formuló. El facilitador aquí debe agrupar los criterios, eliminando las repeticiones literales y reformulando las similares, de manera que todos los expertos vean sus ideas reflejadas. Aquí se deben agrupar los criterios en un orden de aparición, nunca en orden de importancia, dado que el facilitador no debe participar en el proceso de forma directa.

Expertos	Criterio 1	Criterio 2	Criterio n
Experto 1		X		
Experto 2	X			X
.....				
Experto m	X	X		

- **Segunda Ronda.** Los criterios agrupados en la tabla antes expuesta le son enviados a los expertos para que marquen con una X los que consideren mas importantes, aquí van los criterios nunca quién los emitió porque el anonimato debe continuar hasta el final. El facilitador recibe los resultados de los expertos y entonces debe iniciar una labor de eliminación de aquellos que no representan a la mayoría de los expertos. Los criterios que estén avalados por menos del percentil, elegido para la investigación, (por ejemplo los que representen menos del 20 % de los expertos) se eliminan. Algunos trabajan con los intervalos de confianza del 1 %, 5 %, 10 %, o dejar lo que quede dentro del intervalo de confianza (Media- 3 Desviación Estandar , Media + 3 Desviación Estandar). En forma general esta ronda logra hacer la decantación del los criterios minoritarios, si hubiese problemas con esto entonces se debería hacer una ronda adicional para precisar bien los criterios eliminados. Con los criterios finalmente aceptados se pasa a la próxima Ronda.
- **Tercera Ronda.** Esta es la Ronda de la Pesada de los criterios para cada experto. Se envían los criterios definitivamente aceptados a los expertos y se les pide que deben determinar el Peso de cada Criterio en la forma desde el número 1, el criterio mejor o de mayor importancia, hasta el número N, el criterio de menor importancia o último en ser seleccionado. Quiere esto decir que

los expertos deben poner un valor a cada criterio, número entre el 1 y el N, se pueden dar iguales valores a los criterios que los expertos consideren que tienen la misma importancia, a este hecho se le llama LIGADURA. Así el facilitador obtiene una tabla parecida a la antes vista pero con los valores numéricos dados por cada experto a cada criterio:

Expertos	Criterio 1	Criterio 2	Criterio n
Experto 1	R11	R12		R1n
Experto 2	R21	R22		R2n
.....				
Experto m	Rm1	Rm2		Rmn

Donde R_{ij} es la evaluación que da el experto i al criterio j , este valor está en el rango $(1,n)$, n es el número de criterios, m es el número de expertos.

Llega ahora el aspecto estadístico del Método Delphi, que hasta este punto ha sido cualitativo netamente. Se debe calcular el **coeficiente de Kendall** para la prueba de los expertos, que no es más que un coeficiente de regresión lineal que nos da el grado de correlación entre los expertos o la llamada concordancia, este es un índice entre cero y uno $K = 0$ significaría que no existe concordancia entre los expertos, no están de acuerdo con las ideas reflejadas en el trabajo, $K = 1$ significaría que existe concordancia perfecta entre los expertos con los criterios y con el orden de los mismos, si el valor es negativo indicaría que el orden de los criterios no es del acuerdo de todos.

$$K = \frac{\left[12 \sum_{j=1}^n (S_j - S_{med})^2 \right]}{\left[m^2 (n^3 - n) - m \sum_{i=1}^m t_i \right]}$$

donde se tienen las siguientes fórmulas para los elementos anteriores:

S_j es el valor que dan los m expertos al criterio j , S_{med} es el valor medio de los m criterios y T_j es el resultado de rangos iguales o

ligaduras para el experto i donde aquí l es el número de grupos con iguales valores para dicho experto (ligaduras) y t es el número de observaciones dentro de cada uno de los grupos para el experto i (valor de las ligaduras), si todas las evaluaciones del experto i son diferentes no hay ligadura y por lo tanto $T_j = 0$.

$$S_j = \sum_{i=1}^m R_{ij}$$

$$S_{med} = \frac{\sum_{j=1}^n S_j}{n} = \frac{m(n+1)}{2}$$

$$T_j = \sum_{j=1}^l (t^3 - t) / 12$$

Las hipótesis que se plantean con K (Coeficiente de Kendall) son del tipo:

Hipótesis nula: **$H_0 : K = 0$** No hay comunidad de intereses en los expertos con relación a los criterios

Hipótesis alternativa: **$H_1 : K \neq 0$** Los expertos están de acuerdo con los criterios, hay comunidad de intereses.

La Zona de aceptación para que se cumpla la Hipótesis Nula será:

$$\chi^2_{calculada} = m(n-1)K$$

$$\chi^2_{tabulada} \text{ (Nivel de significación, Grados de Libertad)}$$

Si $\chi^2_{calculada} < \chi^2_{tabulada}$ (Nivel de significación, Grados de Libertad) entonces se acepta la Hipótesis nula. Hay comunidad de intereses, en caso contrario, se rechaza la hipótesis nula y se acepta la Hipótesis Alternativa, llegando a concluir que no hay comunidad de intereses entre los expertos.

Ejemplo del Cálculo del Delphi:

Supongamos que un investigador inició un estudio sobre los criterios que tienen diferentes expertos sobre cómo iniciar un tema de investigación. Para ello escogió una lista de expertos de los cuales finalmente, después de hacer el estudio de competencia quedó definitivamente con 3 expertos y luego de tres rondas se tiene la siguiente tabla de resultados que quiere calcular para comprobar las hipótesis de concordancia o no del método.

Expertos m = 3	C1. Necesidad sobre un tema	C2.Hacer estudio previo	C3.Qué exista un problema	C4.Objetivos Determinados	C5. Buena Justificación	C5. Que exista novedad
Experto 1	1	2	1	3	4	4
Experto 2	2	3	1	3	4	5
Experto 3	1	3	2	4	5	3
S_j $\bar{S} = 8.3$	4	8	4	10	13	11

Ligaduras: Experto 1 2 ligaduras { valor 1 $t_1=2$, valor 4 $t_2=2$ }
 Experto 2 1 ligadura { valor 3 $t_1=2$ }
 Experto 3 1 ligadura { valor 3 $t_1=2$ }

$$T_1 = (2^3 - 2)/12 + (2^3 - 2)/12 = 1$$

$$T_2 = (2^3 - 2)/12 = 1/2$$

$$T_3 = (2^3 - 2)/12 + (2^3 - 2)/12 = 1$$

$$K = 12 (83.202) / [9 (216 - 6) - 3 (1 + 1/2 + 1)] = 0.5303$$

$$\chi^2_{\text{calculada}} = m (n - 1) K = 3 (5) 0.5303 = 7.9555$$

$$\chi^2_{\text{tabulada}} (\text{Nivel de significación, Grados de Libertad}) = \chi^2_{\text{tabulada}} (0.05, 5) = 11.07$$

Como $7.9555 \leq 11.07$ Entonces se cumple la Hipótesis Nula H_0 y se puede concluir que no hay comunidad de intereses entre los expertos.

Capítulo 3: Elementos de la Estadística Matemática.

El desarrollo científico e investigativo implica el uso de métodos estadísticos que necesitan el acopio de información o **datos científicos**. La **inferencia estadística** no es sólo acopio de información científica sino un grupo grande de herramientas analíticas que permiten al ingeniero o al científico comprender mejor los sistemas que generan los datos. La información se colecta en forma de **muestras** o agrupación de observaciones. La muestra permite trabajar con menor número de observaciones de la **población** y con su estudio se logran resultados confiables, se facilitan los cálculos y se ahorra dinero.

En un conjunto de datos (**muestra**) de una población se definen algunas de las llamadas **Medidas de Posición** entre las que se encuentran:

Medidas de Posición:

La **Media Aritmética** (o medida de tendencia central) $\bar{X} = \sum_{i=1}^n X_i / n$

Valor promedio aritmético o medida de tendencia central.

La **Mediana** (valor central de la muestra) Si el número de elementos de la muestra n es:

Par, entonces $X_{mediana} = (X_{n/2} + X_{n/2+1})$

Impar, entonces $X_{mediana} = (X_{(n+1)/2})$

La **Moda** (valor más frecuente) $X_{moda} = (\text{Valor que más se repite})$.

Medidas de Variación:

Rango de la Muestra: $X_{max} - X_{min}$

Desviación Cuadrática sobre la media: $S^2 = \sum_{i=1}^n (x_i - \bar{X})^2 / (n-1)$ **o**
varianza.

Grados de Libertad: Número de elementos de información independientes disponibles para calcular la variabilidad (n - 1)

3.1.- Probabilidad.

La sed insaciable del hombre por el juego lo condujo al desarrollo temprano de la Teoría de las Probabilidades, ya en tiempos pasados hablaron sobre el tema Pascal, Leibniz, Fermat, Bernoulli y otros.

La probabilidad es un número positivo entre 0 y 1 que se le asigna a un evento o elemento de un espacio muestral o población. Supongamos un experimento cuyo resultado no se conoce de antemano.

Sea E el espacio muestral o población del experimento, esto es el conjunto de todos los posibles resultados, a los que les llamaremos elementos o sucesos elementales.

Sea $E = \{E_1, E_2, \dots, E_n\}$

E – espacio muestral o población

E_i – suceso elemental o evento

Ejemplo: Suponga que el experimento es una carrera de caballos numerados del 1 al 4.

El espacio muestral lo consideraremos el ordenamiento de los lugares en la carrera. Así por ejemplo si la carrera puede tener un elemento que sea (4,1,3,2) es decir el caballo 4 llegó en primer lugar,....., el caballo 2 llegó en el último lugar.

E = Espacio muestral = {todos los ordenamientos de 1,2,3,y 4} cualquier subconjunto de ese se conoce como suceso elemental o evento.

Si el subconjunto del espacio muestral es:

$S = \{\text{todos los resultados de E que comiencen con 1}\}$

Entonces S significa todos los eventos en que el caballo 1 llegue de primero.

Definición Clásica de Probabilidad.

La probabilidad es un número que se le hace corresponder a cada evento E_i del espacio muestral y que cumple las condiciones:

1. $p(E_i) \in (0,1)$, $0 \leq p(E_i) \leq 1$
2. $p(E) = 1$, la probabilidad del espacio muestral o población es 1.
3. Para cualquier secuencia de eventos mutuamente excluyentes E_1, E_2, \dots se cumple:

$$p(E_1 \cup E_2 \cup E_3 \cup \dots) = \sum p(E_i)$$

Definición Clásica de Probabilidad :

La probabilidad de un suceso o evento S está dada por la relación:

Casos favorables / Casos posibles

Donde los casos posibles significan el conjunto de todos los sucesos que forman esa población y el conjunto de casos favorables es el de aquellos sucesos que satisfacen la condición que se define a priori. En el cálculo de los casos favorables y los posibles intervienen generalmente las fórmulas de las permutaciones, variaciones, combinaciones,... según sea cada caso. También se involucran las operaciones con conjuntos, unión, intersección, complemento, etc.

Elementos de Teoría Combinatoria

Combinaciones de n elementos del espacio muestral tomados en grupos de m elementos: $C_n^m = n! / m!(n-m)!$ significa formar grupos de m elementos del espacio total de n , pero en donde entre cada elemento del grupo intervenga la naturaleza de alguno de sus miembros.

Ejemplo: Si se tiene el conjunto $\{1,2,3\}$ entonces las combinaciones de los 3 elementos en grupos de 2 elementos serán: $C_3^2 = 3! / 2! 1! = 3$ que serán los grupos siguientes: $\{(1,2), (1,3), (2,3)\}$

Variaciones de n elementos del espacio muestral tomados en grupos de m elementos: $V_n^m = n(n-1)\dots(n-m+1)$ significa formar grupos de m elementos del espacio total de n , pero donde intervengan tanto el orden como la naturaleza.

Ejemplo: Si se tiene el conjunto $\{1,2,3\}$ entonces las variaciones de los 3 elementos en grupos de 2 elementos serán : $V_3^2 = 3 \cdot 2 = 6$ que serán los grupos siguientes: $\{(1,2), (2,1), (1,3), (3,1), (2,3), (3,2)\}$.

Ahora si se permite la repetición de un mismo elemento entonces se tiene las **Variaciones con Repetición** $VR_n^m = n^m$ significa que interviene el orden, la naturaleza pero se pueden repetir los mismos elementos.

Ejemplo: Si se tiene el conjunto $\{1,2,3\}$ entonces las variaciones con repetición de los 3 elementos en grupos de 2 con repetición serán: $VR_3^2 = 3 \cdot 3 = 9$ que serán los grupos siguientes $\{(1,2), (2,1), (1,3), (3,1), (2,3), (3,2), (1,1), (2,2), (3,3)\}$.

Permutaciones de m elementos del espacio muestral $P! = m(m-1)(m-2)\dots 1$ cuando se ordenan los m elementos cambiando solamente el orden de alguno de ellos.

Ejemplo: Permutaciones de 3 elementos $\{1,2,3\}$ $P! = 3 \cdot 2 \cdot 1 = 6$
 $\{(1,2,3), (1,3,2), (3,1,2), (3,2,1), (2,3,1), (2,1,3)\}$.

Ejemplos:

1. Una obra de cuatro tomos se coloca en el estante al azar, calculemos la probabilidad de que los tomos queden en orden, ya sea de derecha a izquierda o de izquierda a derecha.

Población = Espacio muestral = conjunto de los casos posibles = Permutaciones de cuatro

$E = P_4 = 4 \cdot 3 \cdot 2 \cdot 1 = 4! = 24$ (factorial de 4 = 24 casos)

Casos favorables = Es el conjunto de ordenamiento en orden directo o inverso de los libros = $\{(1,2,3,4), (4,3,2,1)\}$ 2 casos

Probabilidad = $2/24 = 1/12$

2) Siete personas toman un elevador en un edificio de 10 pisos, calculemos la probabilidad de que todos se bajen en pisos diferentes.

Población = Espacio muestral = Casos Posibles. Todas las posibles formas de descender 7 personas en 10 pisos = Variaciones con repetición de 7 personas en 10 pisos = $VR_7^{10} = 10^7$ elementos

Casos favorables = Variaciones sin repetición de 10 elementos tomados 7 a 7

$$V_{10}^7 = 10!/3! = 7!$$

Luego la probabilidad pedida = $10^7 / 7! = 0.06048$

3.2.- Probabilidad como Función.

Definamos una función cuyo Dominio sea X el conjunto de los posibles resultados que son aleatorios, es decir que suceden según una determinada posibilidad que está relacionada con un valor del intervalo [0,1]. A todos los $X \in X$ se les llama variables aleatorias debido a que representan el resultado de una actividad que es aleatoria. El Codominio de esta función será el intervalo [0,1] y a sus valores le llamaremos la probabilidad o la ley de densidad de la variable aleatoria X.

$$f: X \rightarrow [0,1]$$

f – es la función de probabilidad o de densidad de la variable aleatoria X.

3.3.- Variables Aleatorias.

Una variable aleatoria es una función que asocia un número real con cada elemento del espacio muestral. Un ejemplo de variable aleatoria: Sea el espacio muestral que detalla el resultado de cada uno de los tres componentes electrónicos que se pueden tomar al azar.

$$S = \{ BBB, BBM, BMB, MBB, BMM, MBM, MMB, MMM \}$$

Donde B= Componente en estado Bueno, M= Componente en estado Malo

Asignémosle a cada punto del espacio muestral un valor numérico que represente el # de componentes buenos en el trio con lo que se tendrían los siguientes resultados:

{3, 2, 2, 2, 1, 1, 1, 0}

Las variables aleatorias pueden ser discretas o continuas como ya analizamos anteriormente su definición en forma general:

3.3.1.- Discretas: cuando toma uno de los posibles valores. $X_1, X_2, \dots, X_n, \dots$ y se cumple que:

A cada valor de la variable aleatoria X se le hace corresponder un valor numérico que es su probabilidad $0 \leq p(X_i) \leq 1$, donde $p(X_i)$ es la probabilidad de la variable aleatoria discreta X_i . Es decir que se pueden contar sus resultados posibles. Datos contados como por ejemplo, número de equipos defectuosos, número de accidentes en una vía, número de enfermos en una escuela, etc.

Ejemplo: Sea el Espacio muestral $X = \{1,2,3\}$ con los valores probables de las variables X_1 y X_2 siguientes $p(1) = 1/4$; $p(2) = 1/3$
Luego como $p(1) + p(2) + p(3) = 1$, entonces $p(3) = 5/12$

Una variable aleatoria discreta asume a lo más un conjunto numerable de posibles valores, es frecuente que se analicen variables aleatorias cuyo conjunto de valores sea un intervalo.

Función de probabilidad o de densidad de una variable aleatoria discreta.

Cada variable aleatoria toma para cada uno de sus valores una cierta probabilidad y se le llama función de densidad o de probabilidad al conjunto de pares ordenados $(X, p(X))$ que a cada variable discreta le hace corresponder su probabilidad. Al conjunto de pares ordenados o función de densidad se le imponen las siguientes condiciones:

X variable aleatoria discreta.

$0 < = p(X) < = 1$, su probabilidad o densidad es un valor entre cero y uno.

$\Sigma p (X_i) = 1$, la suma de todas las probabilidades de X_i es igual a uno.

$P (X = k) = p (X)$, valor puntual de la probabilidad de la variable aleatoria X .

Ejercicio: Existen 8 computadoras en un almacén de las que se conoce que 3 de ellas están defectuosas. Si una escuela va a retirar al azar 2 de las computadoras, encuentre la función de probabilidad o densidad de la variable aleatoria número de computadoras defectuosas.

La variable aleatoria $X = \{0, 1, 2\}$

- 0- computadoras defectuosas
- 1- computadoras defectuosas
- 2- computadoras defectuosas

La función de probabilidad o densidad: $p (X = x) = p (X) = \binom{3}{x} \binom{5}{2-x} / \binom{8}{2}$

De 5 computadoras buenas, 3 defectuosas escoger 2.

Así se tienen los valores:

$P (0) = 10/28$ $P (1) = 15/28$ $P (2) = 3/28$ y se tiene la Tabla de la función de probabilidad o densidad.

Valor de X	0	1	2
Probabilidad $p (x)$	10/28	15/28	3/28

Ejercicio: Si una agencia vende autos de su inventario con un 50 % equipado con bolsas de aire, encuentre la función de probabilidad del número de autos con bolsas de aire entre los siguientes 4 autos que vende la agencia.

X.- número de autos con bolsas de aire a elegir = { 0, 1, 2, 3, 4 }

Casos favorables: las combinaciones de autos que se puedan escoger con bolsas de aire, combinaciones de x autos de 4 posibles = $\binom{4}{x}$

Casos posibles: las combinaciones con repetición de 4 autos = $CR_4 = 4^2 = 16$

$$P(0) = 1/16 \quad P(1) = 4/16 \quad P(2) = 6/16 \quad P(3) = 4/16 \quad P(4) = 1/16$$

Tabla de probabilidad o de densidad:

Valor de X	0	1	2	3	4
Probabilidad p (X)	1/16	4/16	6/16	4/16	1/16

Función de distribución o probabilidad acumulada de una variable aleatoria discreta.

E denota por F (X) y es la suma de todas las probabilidades acumuladas hasta el momento.

**F (X) = P (X <= x) = Σ p (Xi) = p (X1)+p(X2)+p(X3)+...p(Xx)
para todos los valores de X <= x**

Para el caso de los ejemplos anteriores se tienen:

Tabla de la Función de Distribución o Acumulada para el ejemplo de las computadoras:

Valor de X	0	1	2
Probabilidad p (x)	10/28	15/28	3/28
P (X <= x)	10/28	25/28	1

Tabla de la Función de Distribución o Acumulada para el ejemplo de los autos con bolsas de aire:

Valor de X	0	1	2	3	4
Probabilidad p (X)	1/16	4/16	6/16	4/16	1/16
P (X <= x)	1/16	5/16	11/16	15/16	1

Es importante para cada variable aleatoria discreta conocer su Valor Esperado y su Varianza.

$$\text{Valor Esperado o Media} = \sum_{i=1}^n XiP(Xi) = \mathbf{E(X)} = \mu$$

$$\text{Valor de la Varianza} = \sum_{i=1}^n (Xi - E(X))^2 p(Xi) = \mathbf{V(X)} = \sigma^2 = \mathbf{E(X^2) - E(X)^2}$$

Ejemplos de Distribuciones Discretas.

a) *Distribución de Bernoulli.*

Consideremos un experimento que consiste en pruebas cada una de las cuales puede tomar valores 0 (fallos) o valores 1 (éxitos), con sus probabilidades correspondientes:

Sea p = probabilidad de un éxito

q = probabilidad de un fallo

La variable aleatoria x que denota el número de éxitos en n pruebas de Bernoulli tiene una función de probabilidad o de densidad:

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, \dots, n$$

0 en otro caso

Es decir,

$$p(\underbrace{EEE\dots E}_x \text{ éxitos}, \underbrace{FFF\dots F}_{n-x} \text{ fallos}) = \binom{n}{x} p^x q^{n-x} \quad (\text{probabilidad de que sucedan } x \text{ éxitos en } n \text{ pruebas})$$

Ejemplo: Una producción manufacturera de chips de computadoras presenta un 2 % de fallos en el proceso. Si se toma una muestra de 50 diariamente. Determine la probabilidad que el proceso tenga que parar por fallos mayores a 2 chips en la muestra.

X: variable Bernoulli p= probabilidad de fallos = 0.02 (en este caso tomados como éxitos por ser lo que se pide) q= probabilidad de no fallos = 0.98

Se pide la probabilidad $p(x) = \binom{50}{x} 0.02^x 0.98^{50-x}$ $x = 0,1,2...50$
 0 en otro caso

$$p(x > 2) = 1 - p(x \leq 2)$$

donde

$$\sum_{x=0}^2 \binom{50}{x} (0.02^x) (0.98^{50-x}) = 0.98^{50} + 50(0.02)(0.98)^{49} + 1225(0.02)^2(0.98)^{48} = 0.92 =$$

$$p(x \leq 2)$$

$P(x > 2) = 0.08$ probabilidad de que la producción pare en un día porque presenta mas de 2 chips con fallos.

Media de Bernoulli: $E(x) = np$

Varianza de Bernoulli: $V(x) = npq$

b) Distribución geométrica

Es una variable aleatoria del tipo Bernoulli que se define como el número de fallos antes de llegar al primer éxito:

$$p(x) = q^{x-1} p \text{ si } x = 1,2,....$$

$$0 \text{ en otro caso.}$$

$$p(x) = P(\underbrace{F F \dots F}_{x-1 \text{ fallos}}, \underbrace{E}_{1 \text{ éxito}})$$

Media de la geométrica = $1/p = E(x)$

Varianza de la geométrica = $q/p^2 = V(x)$

Ejemplo: El 40 % de las impresoras Lazer - Jet son rechazadas en una inspección. Encuentre la probabilidad que el primer equipo aceptado sea el tercero inspeccionado. Se sabe que cada inspección es de Bernoulli con $p = 0.6$ y $q = 0.4$.

$$p(3) = q^{3-1} p^1 = 0.4^2 (0.6) = 0.096 \quad \text{Solamente el 10 \% de las veces el tercer equipo es el aceptado.}$$

3.3.3.- Variables Aleatorias Continuas.

Una variable aleatoria x se dice que es continua si su función de probabilidad o de densidad es continua y el valor de la probabilidad de x viene dado por:

$$F(x) = p(x \leq t) = \int_{-\infty}^x f(t) dt$$

$\int_{-\infty}^{\infty} f(t) dt = 1$ para todos los valores de x

$f(t)$ es la función de probabilidad o densidad entonces $F(x) = p(x \leq t)$ es la función de distribución o función acumulada de la variable aleatoria continua x . Los resultados de las variables aleatorias continuas pueden tomar valores en una escala continua. Datos medidos relacionados con pesos, alturas, temperaturas, distancias, etc.

Media o Valor Esperado de la variable $x = E(x) = \int x * f(t) dt$

Varianza de la variable aleatoria $x = V(x) = \int [(x - E(x))^2] * f(t) dt$

Ejemplo: Si la función de densidad de probabilidad de x está dada por

$$f(x) = \begin{cases} 3x^2 & \text{si } 0 < x < 1 \\ 0 & \text{en caso contrario} \end{cases}$$

Se pide saber cuál será el valor esperado o media de x .

$$E(x) = \int_0^1 x f(x) dx = \int_0^1 x (3x^2) dx = 3/4$$

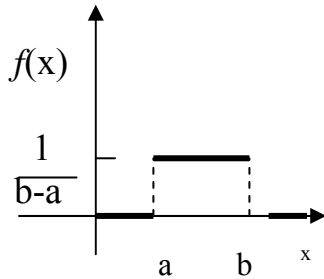
Ejemplos de Variables Aleatorias Continuas:

9. Distribución uniforme.

La distribución uniforme tiene la función de densidad probabilística siguiente:

Una variable aleatoria uniforme es aquella que está uniformemente distribuida en un intervalo (a, b) con un valor de probabilidad igual a $1/(b-a)$ y fuera de este intervalo su valor es 0.

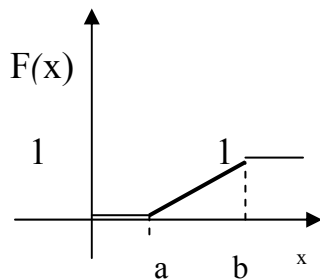
densidad aleatoria $f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{en otro caso} \end{cases}$ función de la variable uniforme.



(la probabilidad de que la variable aleatoria x esté fuera del intervalo $[a,b]$ es cero y dentro del intervalo es 1)

La función de distribución acumulada se expresa como:

$$F(x) = \int_a^x f(x)dx = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a}$$



Media o Valor Esperado de la variable $x = E(x) = (a+b)/2$

Varianza de la variable aleatoria $x = V(x) = (a-b)^2 / 12$

Ejemplo: Suponga que la carga de 3 camiones que siguen una distribución uniforme entre 600 y 700 arrobas. Mediante esta distribución simule la carga que tendrán.

Si x es la variable aleatoria uniforme (carga de un camión en arrobas) entonces:

$$F(x) = (x - a)/(b-a) = (x-600) / 100 \quad \text{donde } a = 600 \text{ y } b = 700$$

Pero como $F(x) \in [0,1]$ generemos 3 números aleatorios en dicho rango y entonces simulemos la carga.

Camión 1 x1. Número aleatorio (de una tabla) = 0,842

Si $F(x_1) = 0,842$, entonces

$$0,842 = (x_1 - 600) / 100 \Rightarrow x_1 = 684.2 \text{ arrobas}$$

Camión 2 x2. Número aleatorio = 0,105

Si $F(x_2) = 0,105$ entonces

$$0,105 = x_2 - 600 / 100 \Rightarrow x_2 = 611.5 \text{ arrobas}$$

Camión 3 Se deja como deber:

Ejemplo: En un proceso productivo se conoce que la producción nunca ha estado por debajo de 2 Ton/día ni más de 4 Ton/día. Se quiere encontrar la distribución de dicha variable aleatoria, su media y su varianza.

$$f(x) = 1/(b - a) = 1/2 \quad \text{si } 2 \leq x \leq 4$$

0 en otro caso

$$F(x) = 0 \quad \text{si } x < 2$$

$$F(x) = (x - a)/(b-a) = (x-2)/2 \quad \text{si } 2 \leq x \leq 4$$

$$F(x) = 1 \quad \text{si } x \geq 4$$

$$E(x) = (a+b)/2 = 3 \text{ Ton / día}$$

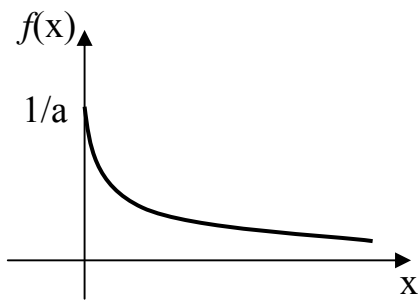
$$V(x) = (a-b)^2 / 12 = 1/3 \text{ Ton / día.}$$

b) Distribución Exponencial

La distribución de densidad probabilística de una variable aleatoria x con distribución exponencial es:

$$f(x) = (1/a) e^{-x/a} \quad \text{para } x \geq 0$$

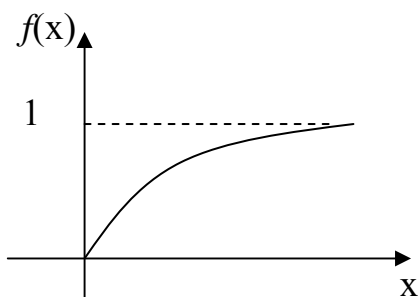
donde a = media de la distribución exponencial



La función de distribución acumulada es:

$$F(x) = \int_0^x \frac{1}{a} e^{-x/a} dx = -e^{-x/a} \Big|_0^x$$

$$= -e^{-x/a} + 1$$



$$F(x) = -e^{-x/9} + 1$$

Se genera una variable aleatoria exponencial de la forma: $x = -a \ln R$ donde R es un número aleatorio buscado de una tabla.

Ejemplo: Simule el tiempo entre llegadas de 2 buques al puerto de Guayaquil si siguen la distribución exponencial con media de 10 horas

Sea $R_1 = 0,3579$ número aleatorio de una tabla

$$x_1 = -10 \ln (0,3579) = 10,27 \text{ horas}$$

(es el tiempo entre llegadas del primer buque)

Sea $R_2 = 0,7349$

$$x_2 = -10 \ln (0,7349) = 3,09 \text{ horas.}$$

Existen otras leyes de distribución que no están al alcance del curso.

Ejercicios:

1) En cierta comunidad, la probabilidad de que una persona de 70 años llegue a los 80 es 0,64.

¿Cuál es la probabilidad de que una persona que tiene 70 años y es miembro de esta comunidad muera dentro de los próximos 10 años?

E- suceso de que una persona de 70 años viva más de 10 años

E¹- complemento, es el suceso de que una persona de 70 años muera dentro de los próximos 10 años.

$$p(E) = 0,64$$

$p(E^1) = 1 - p(E) = 0,36$ es la probabilidad de que una persona de 70 años muera dentro de los próximos 10 años.

2) En una cierta comunidad el 30% de las personas son fumadoras, el 55% son bebedoras y el 20% tanto fumadoras como bebedoras. Si pide encontrar las probabilidades de:

- a) que una persona fuma pero no beba
- b) ni fume ni beba
- c) fume o no beba

Tenemos la tabla siguiente:

	Fuma	No Fuma	Total
Bebe	0,20		0,55
No bebe			
Total	0,30		1

Sea F el suceso de que fume
B el suceso de que tome

$$p(F) = 0,3; \quad p(B) = 0,55$$

$$p(F \cap B) = 0,2$$

Completen la tabla dejando para último la región sombreada

	Fuma	No Fuma	Total
Bebe	0,2	$0,55 - 0,2 = 0,35$	0,55
No bebe	$0,3 - 0,2 = 0,1$	$0,45 - 0,1 = 0,35$ $0,7 - 0,35 = 0,35$	$1 - 0,55 = 0,45$
Total	0,3	$1 - 0,3 = 0,7$	1

Luego

	F	F_1	
B	0,2	0,35	0,55
B^1	0,1	0,35	0,45
	0,3	0,7	1

a) probabilidades de que fume y no beba:

$$p(F \cap B^1) = 0,1$$

b) probabilidad que ni fume ni beba

$$p(F^1 \cap B^1) = 0,35$$

c) probabilidad de que fume o no beba:

$$p(F \cup B^1) = P(F) + P(B^1) - P(F \cap B^1)$$

$$= 0,3 + 0,45 - 0,1$$

$$p(F \cup B^1) = 0,65$$

Distribución Normal.

Una variable aleatoria continua x con media μ y varianza $\sigma^2 > 0$ se le dice que tiene una función densidad normal dada por la fórmula:

$$f(x) = \frac{1}{(\sigma \sqrt{2\pi})} e^{-1/2[(x-\mu)/\sigma]^2} \quad \text{con} \quad -\infty < x < \infty$$

Se dice que la variable aleatoria $x \in N(\mu, \sigma)$. La distribución Normal tiene las siguientes propiedades:

- $\lim_{x \rightarrow -\infty} f(x) = 0$ $\lim_{x \rightarrow \infty} f(x) = 0$ En la medida que se aleja la función del valor medio tanto por la izquierda como por la derecha la probabilidad tiende a cero.
- $f(x+\mu) = f(x-\mu)$ la función es simétrica alrededor de la media μ .
- $\text{Max } f(x) = \mu$ la media es el valor de máxima probabilidad.

La función de Distribución Normal $F(x) = p(X \leq x) = \int_{-\infty}^x f(x) dx$

Se prueba que si $x \in N(\mu, \sigma)$ entonces $Z = (x-\mu)/\sigma \in N(0,1)$

Z: Distribución Normal Estandar, es decir Normal con media 0 y varianza 1. Este valor aparece en las tablas y es muy útil hacer la conversión para poder encontrar las probabilidades normales de las variables aleatorias con cualquier valor de media y varianza.

Nota: Recordar que la Normal tiene el área sombreada en las Tablas siempre desde el punto hacia la izquierda.

Ejercicios:

1. Si se conoce que la variable aleatoria x tiene una distribución normal con media 50 y varianza 9 determine la probabilidad de que la variable aleatoria sea menor de 56.
 $P(x \leq 56) = p(z \leq (x-50)/3) = p(z \leq 2) = 0.9772$
2. Dada una variable aleatoria continua normalmente distribuida con media 50 y desviación estandar 10, encuentre la probabilidad de que x tome un valor entre 45 y 62.

$$\begin{aligned}
P(45 \leq x \leq 62) &= P(z_1 \leq Z \leq z_2) = P(z \leq z_2) - P(z \leq z_1) \\
&= P\left(\frac{45-50}{10} < z < \frac{62-50}{10}\right) = P(-0.5 < z < 1.2) \\
&= P(z < 1.2) - P(z < -0.5) = 0.8849 - 0.3085 = 0.5764
\end{aligned}$$

3. Encuentre la $P(x > 362)$ si $x \in N(300, 2500)$.

$$\begin{aligned}
P(x > 362) &= 1 - P(x \leq 362) = 1 - P\left(z < \frac{362-300}{50}\right) = 1 - 0.8925 \\
&= 0.1075.
\end{aligned}$$

4. Dada la distribución normal con media 40 y desviación estándar 6, encuentre el valor de x que tiene: a) 45 % del área a la izquierda, b) 14 % del área a la derecha.

a) $P(z < -0.13) = 0.45$

$$z = -0.13 = \frac{(x - \mu)}{\sigma} = \frac{(x - 40)}{6} \quad \text{luego } x = 6(-0.13) + 40 = 39.22$$

b) $P(z < 1.08) = 0.86$ $z = 1.08 = \frac{(x - 40)}{6}$ luego $x = 6(1.08) + 40 = x = 46.48$

Capítulo 4: Teoría de la Estimación.

Hemos estudiado la Teoría de la Inferencia Estadística que consiste en aquellos métodos por los que se realizan inferencias o generalizaciones acerca de una población. La tendencia actual es la distinción entre el método clásico de estimación de un parámetro de la población, por medio del cual las inferencias se basan de manera estricta en información que se obtiene de una muestra aleatoria seleccionada de la población y el método bayesiano que utiliza el conocimiento subjetivo previo sobre la distribución de probabilidad de los parámetros desconocidos junto con información que proporcionan los datos de la muestra.

El método clásico de estimación cuenta con dos tipos:

- estimación puntual
- estimación por intervalos

Estimación Puntual.

La **estimación puntual** es la que trata de estimar parámetros de la población ω mediante una estadística ϖ a través de una muestra de tamaño n .

Ya se han visto antes las fórmulas para estimar la media y desviación estandar de la población mediante las fórmulas de la media y desviación estandar muestrales, así como las fórmulas de la mediana y la moda estimadas de una población dada.

Estimadores puntuales de la Media y varianza poblacional:

Media Aritmética $\bar{X} = \sum_{i=1}^n X_i / n$

La **Mediana**:

Si el número de elementos de la muestra n es:

Par, entonces $X_{mediana} = (X_{n/2} + X_{n/2+1})$

Impar, entonces $X_{mediana} = (X_{n+1/2})$

La **Moda** $X_{\text{mod } a} = (\text{Valor que más se repite})$.

La **Varianza** : $S^2 = \sum_{i=1}^n (x_i - \bar{X})^2 / (n-1)$.

El problema aquí es elegir el mejor estimador de los parámetros poblacionales mediante fórmulas establecidas para el cálculo puntual de dichos valores.

Propiedades de los Estimadores.

Sesgo: es equivalente al error, por lo que se desea tener un estimador que tenga un sesgo mínimo o **cero**.

Estimador Insesgado: Se dice que un estadística cualquiera $\bar{\theta}$ es un estimador insesgado del parámetro θ si se cumple que: $E(\bar{\theta}) = \theta$. Lo que significa que el valor esperado del estimador sea igual al parámetro que se pretende estimar.

Ejemplo: Para **muestras pequeñas** $n \leq 30$ se toma el estimador de la media muestral como $\bar{X} = \sum_{i=1}^n X_i / n$ dado que es un estimador insesgado:

$E(\bar{X}) = E[\sum_{i=1}^n X_i / n] = \sum [E(X_i) / n] = n \mu / n = \mu$ por lo que es insesgado.

Ejemplo: Para **muestras grandes** $n > 30$, se toma la fórmula: $\bar{X} = \sum_{i=1}^n X_i / (n-1)$ luego $E(\bar{X}) = \sum_{i=1}^n E(X_i) / (n-1) = n\mu / (n-1) \neq \mu$ estimador sesgado.

Ejemplo: La **varianza muestral** $S^2 = \sum_{i=1}^n (x_i - \bar{X})^2 / (n-1)$ es un estimador insesgado de la varianza.

Eficiencia de un estimador: De todos los estimadores insesgados de un parámetro cualquiera θ se dice mas eficiente al que presenta **menor varianza**.

4.1.- Estimación por Intervalos.

Una estimación por intervalos de un parámetro poblacional θ es un intervalo de la forma $\overline{\theta}_{menor} \leq \theta \leq \overline{\theta}_{mayor}$, donde los valores extremos del intervalo dependen del valor de la estadística θ para una muestra particular y también de la distribución de muestreo de $\overline{\theta}, \overline{\theta}_{menor}, \overline{\theta}_{mayor}$ se encuentran con la condición:

$$p(\overline{\theta}_{menor} \leq \theta \leq \overline{\theta}_{mayor}) = 1 - \alpha \quad 0 \leq \alpha \leq 1 \text{ nivel de confianza}$$

los valores $\overline{\theta}_{menor}, \overline{\theta}_{mayor}$ límites de confianza.

La distribución muestral \bar{X} está distribuida normal con media $\mu = \mu_X$ y desviación estandar $\sigma = \sigma / \sqrt{n}$

$$P(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) = 1 - \alpha \quad \text{con } Z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$$

$$P(-Z_{\alpha/2} \leq (\bar{x} - \mu) / (\sigma / \sqrt{n}) \leq Z_{\alpha/2}) = 1 - \alpha$$

Despejando $p(\bar{x} - Z_{\alpha/2} \sigma / \sqrt{n} \leq \mu \leq \bar{x} + Z_{\alpha/2} \sigma / \sqrt{n}) = 1 - \alpha$ Intervalo de confianza de la media poblacional al $(1 - \alpha) 100\%$.

4.1.1.- Estimación por Intervalos de la media poblacional cuando la Varianza es conocida.

Si \bar{X} es la media muestral de un muestreo aleatorio de tamaño n de una población con varianza σ^2 , conocida, un intervalo de confianza de $(1 - \alpha) 100\%$ para la media μ está dado por :

$$\bar{x} - Z_{\alpha/2} \sigma / \sqrt{n} \leq \mu \leq \bar{x} + Z_{\alpha/2} \sigma / \sqrt{n}$$

donde $Z_{\alpha/2}$ es el valor que deja un área de $\alpha/2$ a la derecha.

Si $n \geq 30$ el resultado está garantizado.

Ejemplo: Se encuentra que la concentración promedio de zinc que se saca del agua a partir de una muestra de mediciones de zinc en 36 sitios diferentes es de 2.6 gramos por mililitro. Encuentre los intervalos de confianza de 95 % y 99 % para la concentración media de zinc en el río. Se conoce que $\sigma = 0.3$ de resultados anteriores.

Aquí se tiene que:

$$\bar{X} = 2.6 \text{ gramos}$$

$$n = 36 \text{ sitios de mediciones}$$

$$\sigma = 0.3$$

$$\text{Si } 1-\alpha = 0.95$$

$$\text{Calculando: } Z_{0.925} = 1.96$$

$$Z_{0.025} = -1.96$$

$$\sigma/\sqrt{n} = 0.3/6$$

se obtiene el intervalo de confianza para la media poblacional, aplicando la fórmula anterior: $2.50 \leq \mu \leq 2.70$ gramos.

$$\text{Si } 1-\alpha = 0.99$$

$$\text{Calculando: } Z_{0.995} = 2.575 \quad \text{el intervalo de confianza ahora será:}$$

$$Z_{0.005} = -2.575 \quad 2.47 \leq \mu \leq 2.47 \quad \text{gramos}$$

4.1.2.- Estimación por Intervalos de la media poblacional cuando la varianza es desconocida:

Ahora no se conoce la varianza por lo que se tiene que trabajar con el estimador de la misma s y la fórmula será con la utilización de la distribución **t Student** en lugar de la Normal:

$$\bar{X} - t_{\alpha/2}(n-1)\sigma / \sqrt{n} \leq \mu \leq \bar{X} + t_{\alpha/2}(n-1)\sigma / \sqrt{n}$$

Notas sobre la Distribución t de Student: La distribución t es una variante de la Normal en donde se supone que haya un mejor comportamiento alrededor de la media para variaciones de la varianza, es una distribución simétrica, aunque en la mayoría de los textos los valores que aparecen en las tablas son el área a la derecha del valor, no el de la

izquierda como en la normal, también interviene el grado de libertad (n-1) de la distribución para corregir errores. Si se encuentra el valor de $t_{\alpha}(n-1)$ a la derecha entonces se tendrá lo siguiente: para el simétrico $t_{1-\alpha}(n-1)$ sólo se cambiará el signo. $t_{\alpha}(n-1) = - t_{1-\alpha}(n-1)$

4.1.4.- Intervalos de confianza para la Varianza σ^2 de una población normalmente distribuida.

En este caso se trabajará con la distribución χ^2 y la fórmula será :

$$(n-1) s^2 / \chi^2_{\alpha/2} (n-1) \leq \sigma^2 \leq (n-1) s^2 / \chi^2_{1-\alpha/2} (n-1)$$

La distribución χ^2 no es simétrica por lo que hay que buscar sus ambos valores a la izquierda y a la derecha en forma indistintiva y tiene la característica que se dan sus valores también en la cola derecha, es decir hacia la derecha del punto. Si se quiere buscar un intervalo de confianza para el nivel de significación del $\alpha = 0.05$, entonces se tendrán que buscar los valores siguientes:

A la derecha: $\chi^2_{\alpha}(n-1)$

A la izquierda: $\chi^2_{1-\alpha/2}(n-1)$

Ejemplo:

a).- Suponga una muestra de 20 elementos y se quiere encontrar los límites de la distribución χ^2 al 95 % de confiabilidad

b).- Encuentre el intervalo de confianza para la varianza si se sabe que la varianza muestral es de 1.20^{-4}

a)

$$n = 20$$

$$1-\alpha/2 = 0.975$$

$$\alpha/2 = 0.025$$

$$\text{Extremo derecho: } \chi^2(\alpha/2, n-1) = 32.852$$

$$\text{Extremo izquierdo: } \chi^2(1-\alpha/2, n-1) = 8.907$$

b).- Intervalo de Confianza para σ

$$(n-1) s^2 / \chi^2_{\alpha/2} (n-1) \leq \sigma^2 \leq (n-1) s^2 / \chi^2_{1-\alpha/2} (n-1) \text{ Sustituyendo}$$

$$8.33^{-3} < \sigma < 1.599^{-2}$$

Ejemplo:

Los pesos en decagramos de 10 paquetes de semillas de pasto distribuidas en una zona son:

46.6; 46.1; 45.8; 47; 46.1; 45.9; 45.8; 46.9; 45.2; 46

Encuentre un intervalo de confianza al 95 % para la varianza de los paquetes de semillas, si éstos tienen una distribución normal.

$$S^2 = [n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2] / n(n-1) = 0.286$$

$$\alpha = 0.05 \quad n-1 = 9 \text{ grados de libertad} \quad \chi^2_{(0.025, 9)} = 19.023 \quad \chi^2_{(0.075, 9)} = 2.7$$

$$0.135 < \sigma^2 < 0.953$$

Capítulo 5.- Pruebas de Hipótesis

Ya anteriormente estudiamos las hipótesis en la investigación, ahora comenzaremos el estudio de las hipótesis estadísticas, una hipótesis estadística es una aseveración o conjetura con respecto a una o mas poblaciones. Las hipótesis se pueden aceptar o negar, una hipótesis se acepta simplemente cuando los datos no dan suficientemente evidencia para rechazarlas. Por otro lado, el rechazo implica que la evidencia muestral la refuta.

Hipótesis nula: es la hipótesis que deseamos probar. H_0

Hipótesis alternativa: es la hipótesis que se obtiene cuando se rechaza la hipótesis nula. H_1

En el Proceso de las hipótesis existen dos tipos de errores:

- Rechazar H_0 si es verdadera.- **ERROR DEL TIPO 1.** Este error se puede controlar mediante el Nivel de Significación: α .
- Aceptar H_0 si ésta es falsa.- **ERROR DEL TIPO 2.** Este error se analiza mediante la potencia: β .

Prueba de Hipótesis	H_0 es Verdadera	H_0 es Falsa
Aceptar H_0	Decisión Correcta	Error tipo 2
Rechazar H_0	Error tipo 1	Decisión Correcta

5.1.- Prueba de Hipótesis de una muestra para la media con varianza poblacional conocida. $N(0,1)$.

Sea $X = \{x_1, x_2, x_3, \dots, x_n\}$ es una muestra de tamaño n de una población que tiene media μ desconocida y varianza conocida σ^2 , las hipótesis, para un nivel de significación α , serán :

➤ Prueba Bilateral.

$H_0 : \mu = \mu_0$ Hipótesis Nula
 $H_1 : \mu \neq \mu_0$ Hipótesis Alternativa

$$Z = (\bar{x} - \mu) / (\sigma / \sqrt{n}) \qquad Z_{\text{calc}} = (\bar{x} - \mu_0) / (\sigma / \sqrt{n})$$

Zona de Aceptación: $-Z_{\alpha/2} \leq Z_{\text{calc}} \leq Z_{\alpha/2}$

Zona de Rechazo: $Z_{\text{calc}} > Z_{\alpha/2} \quad \text{o} \quad Z_{\text{calc}} < -Z_{\alpha/2}$

Si se quiere escribir la región crítica en términos de la media muestral

$X > b$ o $X < a$ donde $b = \mu_0 - Z_{\alpha/2} (\sigma / \sqrt{n})$ $a = \mu_0 + Z_{\alpha/2} (\sigma / \sqrt{n})$

➤ **Prueba Unilateral.**

1. $H_0 : \mu = \mu_0$ Hipótesis Nula
 $H_1 : \mu > \mu_0$ Hipótesis Alternativa

Zona de Aceptación: $Z_{\text{calc}} \leq Z_{\alpha}$
Zona de Rechazo: $Z_{\text{calc}} > Z_{\alpha}$

2. $H_0 : \mu = \mu_0$ Hipótesis Nula
 $H_1 : \mu < \mu_0$ Hipótesis Alternativa

Zona de Aceptación: $Z_{\text{calc}} \geq Z_{\alpha}$
Zona de Rechazo: $Z_{\text{calc}} < -Z_{\alpha}$

Ejemplo: Se estudia el promedio de vida en una región dada del país. En una muestra de 100 elementos realizada se obtuvo un promedio de vida de 71.8 años. Si la desviación estándar es de 8.9 años, se puede decir que la media de vida hoy es mayor a 70 años con un nivel de significación del 95 %.

- $H_0 : \mu = 70$ años Hipótesis Nula
 $H_1 : \mu > 70$ años Hipótesis Alternativa

$$Z_{\text{calc}} = (71.8 - 70) / (8.9 / \sqrt{100}) = 2.02$$

Si $\alpha = 0.05$ $Z_{\alpha} = Z_{0.05} = 1.645$

Como $Z_{\text{calc}} > Z_{0.05}$ Estamos en presencia de la zona de rechazo por lo que se puede asegurar estadísticamente que la vida media hoy es mayor a los 70 años.

Se puede decir también: $p (Z > 0.02) = 1 - p (Z \leq 2.02) = 0.0217$ el nivel de significación para la Hipótesis Alternativa es mas fuerte que el nivel de significación 0.05.

Ejemplo: Un fabricante de un sedal sintético afirma que, el mismo, tiene una resistencia media a la tensión de 8 Kg, con una desviación estándar de 0.5 Kg. Pruebe la hipótesis del fabricante contra la de los técnicos de la empresa que dicen que no es de 8 Kg. Se realizó una muestra de 50 sedales y se encontró una tensión de 7.8 Kg. Utilice el nivel de significación del 99 % para contrastar las hipótesis.

Ho : $\mu = 8$ Kg Hipótesis Nula
 H₁ : $\mu \neq 8$ Kg Hipótesis Alternativa

$\alpha = 0.01$, luego $Z_{\alpha/2} = Z_{0.005} = 2.575$

$Z_{\text{calc}} = (7.8 - 8) / (0.5 / \sqrt{50}) = - 2.83$

$Z_{\text{calc}} = - 2.83 < - Z_{\alpha/2} = - 2.575$ Se está en la Zona de Rechazo de Ho
 La resistencia del sedal no es de 8 Kg.

- **La Prueba de Hipótesis Ho : $\mu = \mu_0$ contra la alternativa H₁ : $\mu \neq \mu_0$ a un nivel de significación α es equivalente a calcular un intervalo de confianza del 100 (1- α) % sobre la media y se rechaza H₀ si no se está dentro del intervalo de confianza.**

Zona de Aceptación:

Bilateral. $\bar{X} - Z\alpha/2(\sigma/\sqrt{n}) < \mu_0 < \bar{X} + Z\alpha/2(\sigma/\sqrt{n})$

Unilateral: $\mu > \mu_0$ $\mu_0 < \bar{X} - Z\alpha(\sigma/\sqrt{n})$

$\mu < \mu_0$ $\mu_0 > \bar{X} + Z\alpha(\sigma/\sqrt{n})$

5.2.- Prueba de Hipótesis de una muestra para la media con varianza poblacional desconocida.

En este caso al tratar con varianza desconocida se debe utilizar la ley de distribución **t de Student**, con el nivel de significación α y $(n-1)$ grados de libertad, en lugar de la Distribución Normal. Se asume el valor s^2 de la varianza muestral.

➤ Prueba Bilateral.

$$\begin{aligned} H_0 : \mu &= \mu_0 && \text{Hipótesis Nula} \\ H_1 : \mu &\neq \mu_0 && \text{Hipótesis Alternativa} \end{aligned}$$

$$t_{\text{calc}} = (X - \mu_0) / s\sqrt{n}$$

$$\text{Zona de Aceptación:} \quad -t_{\alpha/2, n-1} = t_{\text{calc}} < t_{\alpha/2, n-1}$$

$$\text{Zona de Rechazo:} \quad t_{\text{calc}} < -t_{\alpha/2, n-1} \quad \text{O} \quad t_{\text{calc}} > t_{\alpha/2, n-1}$$

➤ Prueba Unilateral:

$$\begin{aligned} 1. \quad H_0 : \mu &= \mu_0 && \text{Hipótesis Nula} \\ H_1 : \mu &> \mu_0 && \text{Hipótesis Alternativa} \end{aligned}$$

$$\text{Zona de Aceptación} \quad t_{\text{calc}} \leq t_{\alpha, n-1}$$

$$\text{Zona de Rechazo} \quad t_{\text{calc}} > t_{\alpha, n-1}$$

$$\begin{aligned} 2. \quad H_0 : \mu &= \mu_0 && \text{Hipótesis Nula} \\ H_1 : \mu &< \mu_0 && \text{Hipótesis Alternativa} \end{aligned}$$

$$\text{Zona de Aceptación:} \quad t_{\text{calc}} \geq -t_{\alpha, n-1}$$

$$\text{Zona de Rechazo} \quad t_{\text{calc}} < -t_{\alpha, n-1}$$

Ejemplo: Un estudio de equipos electrodomésticos afirman que un equipo cualquiera gasta como promedio 46 Kw/h al año. Si una muestra aleatoria de 12 hogares indicó un promedio de 42 Kw/h con una desviación estándar de 11.9 Kw/h ¿ Esto sugiere que en un 0.05 de significación los equipos gasten menos de 46 Kw/h como promedio ?.

$$\begin{aligned} H_0 : \mu &= 46 && \text{Hipótesis Nula} \\ H_1 : \mu &< 46 && \text{Hipótesis Alternativa} \end{aligned}$$

$$\alpha = 0.05$$

$$n = 12$$

$$-t_{0.05, 11} = -1.796$$

$$t_{\text{calc}} = (42 - 46) / (11.9 / \sqrt{12}) = -1.16$$

Zona de Aceptación: $t_{\text{calc}} \geq t_{\alpha, n-1}$ $-1.16 > -1.796$ Se acepta H_0

Zona de Rechazo $t_{\text{calc}} < -t_{\alpha, n-1}$ -1.16 no es menor que -1.796 Se rechaza H_0

Ahora $p(t < -1.16) = 0.135$ es una probabilidad mayor al 5 % por lo que evidencia la zona H_0 .

5.3.- Prueba de Hipótesis para DOS MUESTRA.

1) Se tienen ahora dos **muestras independientes** seleccionadas de dos poblaciones P_1 y P_2 con medias μ_1 y μ_2 y **varianzas conocidas** σ_1^2 y σ_2^2 , se quiere hacer contrastes con ellas, sean éstas:

$$\mathbf{X}_1 = \{ \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n1} \} \quad \text{y} \quad \mathbf{X}_2 = \{ \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n2} \}$$

- Si las **varianzas son desiguales** $[\sigma_1^2 \neq \sigma_2^2]$

Se calcula $\mathbf{Z}_{\text{calc}} = ((\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)) / \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} \notin N(0,1)$

- Si las **varianzas son iguales** $[\sigma_1^2 = \sigma_2^2 = \sigma^2]$

Se calcula $\mathbf{Z}_{\text{calc}} = ((\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)) / \sigma \sqrt{1/n_1 + 1/n_2} \notin N(0,1)$

Pruebas de Hipótesis Bilaterales:

$$\begin{aligned} H_0: \mu_1 - \mu_2 &= d_0 & \text{Zona de Aceptación: } & -Z_{\alpha/2} < Z_{\text{calc}} < Z_{\alpha/2} \\ H_1: \mu_1 - \mu_2 &\neq d_0 & \text{Zona de Rechazo: } & Z_{\text{calc}} > Z_{\alpha/2} \quad \text{o} \quad Z_{\text{calc}} < -Z_{\alpha/2} \end{aligned}$$

Pruebas de Hipótesis Unilaterales:

$$\begin{aligned} \text{a) } H_0: \mu_1 - \mu_2 &= d_0 & \text{Zona de Aceptación: } & Z_{\text{calc}} \leq Z_{\alpha} \\ H_1: \mu_1 - \mu_2 &> d_0 & \text{Zona de Rechazo: } & Z_{\text{calc}} > Z_{\alpha} \end{aligned}$$

$$\begin{aligned} \text{b) } H_0: \mu_1 - \mu_2 &= d_0 & \text{Zona de Aceptación: } & Z_{\text{calc}} \geq -Z_{\alpha} \\ H_1: \mu_1 - \mu_2 &< d_0 & \text{Zona de Rechazo: } & Z_{\text{calc}} < -Z_{\alpha} \end{aligned}$$

2) Si las **varianzas son desconocidas pero iguales**: $\sigma_1^2 = \sigma_2^2$

Utiliza en este caso la **distribución t de student**, la prueba es para la comparación de medias de las dos poblaciones y el valor para la comparación es llamado **d₀**, los estadígrafos son los siguientes:

$$t_{\text{calc}} = ((\bar{X}_1 - \bar{X}_2) - d_0) / (s_p \sqrt{1/n_1 + 1/n_2})$$

donde

$$s_p^2 = (s_1^2(n_1 - 1) + s_2^2(n_2 - 1)) / (n_1 + n_2 - 2)$$

Pruebas de Hipótesis Bilaterales:

$$\begin{aligned} H_0: \mu_1 - \mu_2 &= d_0 \\ H_1: \mu_1 - \mu_2 &\neq d_0 \end{aligned}$$

$$\text{Zona de Aceptación: } -t_{\alpha/2, n_1+n_2-2} < t_{\text{calc}} < t_{\alpha/2, n_1+n_2-2}$$

$$\text{Zona de Rechazo: } t_{\text{calc}} < -t_{\alpha/2, n_1+n_2-2} \quad \text{o} \quad t_{\text{calc}} > t_{\alpha/2, n_1+n_2-2}$$

Pruebas de Hipótesis Unilaterales:

$$\begin{aligned} \text{a) } H_0: \mu_1 - \mu_2 &= d_0 \\ H_1: \mu_1 - \mu_2 &> d_0 \end{aligned}$$

Zona de Aceptación: $t_{calc} \leq t_{\alpha, n_1 + n_2 - 2}$

Zona de Rechazo: $t_{calc} > t_{\alpha, n_1 + n_2 - 2}$

b) **H₀:** $\mu_1 - \mu_2 = d_0$

H₁: $\mu_1 - \mu_2 < d_0$

Zona de Aceptación: $t_{calc} \geq t_{\alpha, n_1 + n_2 - 2}$

Zona de Rechazo: $t_{calc} < -t_{\alpha, n_1 + n_2 - 2}$

Ejemplo: Un experimento para comparar el desgaste por abrasivo de 2 diferentes materiales toma 12 piezas del material 1 y 10 piezas del material 2 en una máquina para probar el desgaste. La muestra 1 da un desgaste medio de 85 unidades y una desviación estándar de 4, la muestra 2 da un desgaste medio de 81 y desviación estándar de 5. ¿A un nivel de 5 % de significación se puede decir que el desgaste del material 1 excede al del material 2 en más de 2 unidades?. Se plantean poblaciones normales con varianzas iguales.

H₀: $\mu_1 - \mu_2 = 2$ $\bar{X}_1 = 85$ $s_1 = 4$ $n_1 = 12$ Varianzas desconocidas pero desiguales

H₁: $\mu_1 - \mu_2 > 2$ $\bar{X}_2 = 81$ $s_2 = 5$ $n_2 = 10$ $\alpha = 0.05$

$t_{calc} = [(85-81)-2] / 4.478 (1/12+1/10)^{1/2} = 1.04$

$s_p = \{ [11*14 + 9*25] / [12+10-2] \}^{1/2} = 4.478$

$t_{\alpha, so} = 1.725$

Zona de Aceptación: $t_{calc} = 1.04 < t_{0.05, 20} = 1.725$ Se cumple hipótesis nula.

No hay razones para decir que el material 2 excede en 2 al material 1 en cuanto al desgaste.

$p(t > 1.04) = 0.16$

Varianzas Desconocidas y diferentes: $\sigma_1^2 \neq \sigma_2^2$ (desconocidas)

$$T_{\text{calc}} = ((\bar{X}_1 - \bar{X}_2) - d_0) / \sqrt{s_1^2 / n_1 + s_2^2 / n_2}$$

Se busca la distribución **t de student** con $\alpha/2$ de nivel de significación y los grados de libertad obtenidos de la fórmula:

$$v = (s_1^2 / n_1 + s_2^2 / n_2)^2 / ((s_1^2 / n_1)^2 / (n_1 - 1) + (s_2^2 / n_2)^2 / (n_2 - 1))$$

Prueba de Hipótesis bilateral:

H₀: $\mu_1 - \mu_2 = d_0$

H₁: $\mu_1 - \mu_2 \neq d_0$

Zona de Aceptación: $-t_{\alpha/2, v} < t_{\text{calc}} < t_{\alpha/2, v}$

Zona de Rechazo: $t_{\text{calc}} < -t_{\alpha/2, v} \quad \text{o} \quad t_{\text{calc}} > t_{\alpha/2, v}$

Distribuciones Pareadas: (con varianzas desconocidas)

Aquí se supone que se tiene dos muestras diferentes, del mismo número de elementos, provenientes de dos poblaciones normales y que los valores se toman en forma pareada:

Nro. del elemento i	Valores Muestra X _{1,i}	Valores Muestra X _{2,i}	Diferencias d _i = X _{1,i} -X _{2,i}
1			
.....			
n			

Media de las diferencias entre las muestras pareadas:

$$D_{\text{media}} = \sum d_i / n$$

Se utiliza la distribución t de student con n-1 grados de libertad.

$$t_{\text{calc}} = (D_{\text{media}} - d_0) / (sd / n^{1/2}) \quad \text{donde se tiene que:}$$

$$sd^2 = (\sum d_i - n D_{\text{media}}^2) / n-1$$

$$H_0 : \mu_D = d_0$$

$$H_1 : \mu_D \neq d_0$$

$$\text{Zona de Aceptación:} \quad - t_{\alpha / 2, n-1} < t_{calc} < t_{\alpha / 2, n-1}$$

$$\text{Zona de Rechazo:} \quad t_{calc} < -t_{\alpha / 2, n-1} \quad \text{o} \quad t_{calc} > t_{\alpha / 2, n-1}$$

Capítulo 6.- Elementos de Muestreo.

Ideas Preliminares

Gran parte de la inferencia estadística se ocupa de problemas en los que se realizan afirmaciones sobre una población a partir de la información de una muestra se tienen dos problemas fundamentales: 1).- Se sabe poco sobre Cómo se seleccionan realmente los elementos de una muestra. 2).- Se asume normalmente que el número de elementos de la población es muy grande comparado con el tamaño de la muestra.

Son Ejemplos de muestreo :

- Las aplicaciones económicas.
- Investigaciones de mercado en poblaciones humanas.
- Las auditorias en las empresas, muestrean las facturas o las cuentas para hacer de ellas inferencias a la población.
- Problemas de los recursos humanos en las empresas en donde se hace un muestreo para saber el cumplimiento de nuevos métodos de producción.
- Sondeos de Opinión: en elecciones sobre los candidatos, en escuelas sobre profesores, alumnos, grupos, facultades, años, grupos y especialidades, entre otros, por parte de la dirección.
- La aplicación de instrumentos en diferentes temas de investigación.
- Otras.

¿Por qué Muestrear?

La alternativa es intentar obtener información de cada elemento de la población. Esto se conoce como CENSO en lugar de muestreo. Hay tres razones importantes por las cuales es mejor aplicar una Muestra y no un Censo:

- Realizar un Censo completo sería muy costoso o prohibitivo.

- Muchas veces se necesita obtener la información con rapidez, un censo completo sería muy demorado y costoso, se podrían obtener resultados cuando ya no sea necesario.
- Con los métodos estadísticos modernos generalmente es posible obtener resultados con el nivel de precisión deseado mediante el muestreo. El tiempo y el dinero que se gaste en producir resultados cuya precisión supere las necesidades del investigador estaría mejor gastado en cualquier otra cosa.

Los pasos a seguir en el muestreo son:

- 1.- Motivación y punto de partida del estudio.
- 2.- Conocimientos de la Población en la Investigación. En función de la lista de la población se puede definir el tipo de muestreo.
- 3.- La muestra seleccionada será obtenida según el Tipo de Muestreo que sea de interés del investigador.
- 4.- Se tratará de: Primero obtener respuestas de una proporción lo más grande posible de miembros de la muestra. Segundo deseando obtener respuestas sinceras. Todo depende del Diseño del Cuestionario.
- 5.- Aplicación de Métodos de inferencias para diseños de muestreo de particulares.
- 6.- Etapa de resumir y presentar la información conseguida, esto puede incluir estimaciones puntuales o por intervalo, tablas, gráficas, etc.

Una vez que se tiene concebida la idea de investigación, que tenemos claro el problema que vamos a estudiar, que se han planteado los objetivos a alcanzar y las hipótesis y preguntas de investigación, que hemos diseñado el tipo de investigación, el siguiente aspecto que debemos definir es la selección del objeto de estudio, o sea establecer la población y la muestra.

En esta parte de la investigación debemos definir quiénes y qué características deberán tener los sujetos (personas, organizaciones o situaciones y factores) objetos de nuestro estudio.

Una tarea importante del investigador es definir cuidadosa y completamente la población antes de recolectar la muestra.

- **Elemento:** Un elemento es un objeto en el cual se toman las mediciones.
- **Concepto de población:** Por población o universo definimos la totalidad de elementos o individuos que poseen la característica que estamos estudiando. Esta población inicial que se desea investigar es lo que se denomina población objetivo. *Población es una colección de elementos acerca de los cuales deseamos hacer alguna inferencia.*
- La población no siempre es posible estudiarla por lo tanto es necesario determinar la muestra a estudiar.
- **Concepto de muestra:** Es cualquier subconjunto de la población que se realiza para estudiar las características en la totalidad de la población, partiendo de una fracción de la población.
- De la muestra es de la que se obtiene la información para el desarrollo del estudio y sobre la cual se efectuarán la medición y la observación de las variables de la investigación a realizarse. *Una muestra es una colección de unidades seleccionadas de un marco o de varios marcos.*
- **Marco:** Un marco es una lista de unidades de muestreo.
- **Unidades de Muestreo:** Son colecciones no traslapadas de elementos de la población que cubren la cubren completamente.

6.1.- Diferentes Tipos de Muestreo.

Podemos dividirlo en dos grandes ramas: **muestreo probabilístico y muestreo no probabilístico.**

➤ Muestreo probabilístico

La Teoría del Muestreo trata el problema de la selección de las muestras adecuadas para lograr el objetivo de obtener la información deseada acerca de la población en estudio.

En este tipo de muestreo todos los elementos tienen la misma probabilidad de ser elegidos.

Para realizar un muestreo probabilístico necesitamos determinar el tamaño de la muestra (n) y seleccionar los elementos muestrales, de manera que todos tengan la misma posibilidad de ser elegidos.

Dentro del muestreo probabilístico existen varias formas de muestreo, de acuerdo al problema que se quiera estudiar:

6.2.- Muestreo aleatorio simple o al azar

La forma más común para obtener una muestra representativa es la selección al azar (aleatoria). **Básico**. Para tener la seguridad de que cada uno de los individuos de la población tienen la misma posibilidad de ser elegidos debe emplearse una **tabla de números aleatorios**, que se encuentran en los libros de estadística. Consiste en seleccionar un grupo de **n unidades muestrales** de tal manera que cada muestra de tamaño n tenga la misma oportunidad de ser seleccionada. Se conoce este muestreo como **MIA** o **MAS** según los autores que sean estudiados. Algunas fórmulas de este tipo de muestreo son:

N tamaño de la población.

n tamaño de la muestra.

Los parámetros poblacionales de media y desviación estándar son denotadas por μ y σ respectivamente.

Los parámetros muestrales de la media y la desviación estándar son \bar{X} y S respectivamente, sus fórmulas ya han sido estudiadas anteriormente.

El estimador de la varianza sobre la media muestral

$V^2 = s^2(N-n)/nN$ donde $(N-n)/N$ es llamado el factor de corrección.

Tamaño de la Población: $T = N \mu$

Estimador del Total de la Población: $T_{\text{estim}} = N \bar{X}$

Número de Elementos de la Muestra.

a) Primeramente, para el caso que no se sepa nada acerca de la población, podemos fijar un valor cualquiera para el número de elementos de la muestra n' , encontraremos el tamaño de la muestra sin ajustarlo a la media y varianzas poblacionales mediante la fórmula:

$$n_1 = \frac{n'}{1 + \frac{n'}{N}}$$

Donde:

N tamaño de la población

n_1 tamaño de la muestra.

n' tamaño provisional de la muestra.

Con este valor de la muestra calculado encontramos la media \bar{X} y la varianza s^2 muestrales y de aquí pasaremos a aplicar la fórmula que se da a continuación, que es la verdadera.

b) Para el caso en que se conozcan el valor la varianza poblacional o en su lugar la varianza muestral, por algún trabajo de investigación realizado con anterioridad, entonces el número de elementos de la muestra se calculará según la fórmula:

➤ Para el caso de varianza poblacional conocida.

$$n = \frac{N \sigma^2}{(N - 1) B^2 / 4 + \sigma^2}$$

- Para el caso de varianza poblacional desconocida y sustituida por la varianza muestral.

$$n = \frac{Ns^2}{(N-1)B^2/4 + s^2}$$

Donde:

N = Tamaño de la población.

σ = Desviación Estándar poblacional.

S = Desviación Estándar de la muestra (conocida o estimada a partir de estudios anteriores o mediante una prueba piloto).

B = Error o diferencia máxima entre la media muestral y la media poblacional que se está dispuesto a aceptar con el nivel de significación escogido. *Este error el propio investigador lo fija.*

Nota: Para el caso del inciso a) en donde no se conocen parámetro de la población de experimentos anteriores, entonces se procede de la forma siguiente:

- Se elige un valor n_1 se calcula con la fórmula dada en a) n_1 .
- Con este valor de n_1 calculado se obtiene una muestra y con ella el valor de la desviación estándar muestral, se aplica la fórmula dada en b) para n.
- Si $n_1 > n$ se trabaja con los valores de media y desviación estándar calculados con n_1 , se tiene una muestra mayor a la necesaria pero es significativa.
- Si $n_1 < n$ entonces se tiene que ir a completar la muestra con los $n-n_1$ elementos restantes para que la muestra sea significativa.

c) En ocasiones se tiene en la población de tamaño N una proporción de elementos que sabemos que cumple con la condición que se analiza digamos:

p proporción de elementos que cumplen la condición.
q = 1- p proporción de elementos que no cumplen la condición.
D = B²/4 donde **B** = Error dado por el investigador.
N = Tamaño de la población.

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

6.3. Muestreo estratificado

Se dice que una muestra es estratificada cuando los elementos de ella son proporcionales a su presencia en la población. La presencia de un elemento en un estrato excluye su presencia en otro, es un requisito del muestreo estratificado que los elementos en cada estrato sean mutuamente excluyentes. En la muestra estratificada la representación de los elementos en los estratos es proporcional a su representación en la población.

Proceso de dos pasos en el que la población se divide en estratos. Los estratos deben ser de forma que a cada elemento de la población le corresponda un solo estrato.

Los elementos se seleccionan de cada estrato por medio de un procedimiento aleatorio, por lo general a través de un muestreo aleatorio simple. Los elementos en un estrato deben ser lo más homogéneos posibles.

Es una técnica muy popular por su gran precisión.

La muestra aleatoria estratificada es la obtenida mediante la separación de los elementos de la población en grupos que no presentan traslapes, llamados estratos, y la selección posterior de una muestra aleatoria simple.

Motivos para la utilización del muestreo aleatorio estratificado:

- El límite de error de estimación es menor, siempre que las mediciones dentro de los estratos sean homogéneas.
- El costo por observación de la encuesta por estratificación se puede reducir.
- Se pueden obtener estimaciones de parámetros poblacionales para subgrupos de la población o estratos.

Fórmulas para el Muestreo Estratificado:

L = Número de estratos.

N_i = Número de unidades totales en el estrato i .

N = Número de unidades de la Población = $N_1 + N_2 + \dots + N_L$

$\gamma_{media\ i}$ = media para el estrato i

γ_{media} = Estimador de la media poblacional μ

$$\gamma_{media} = 1 / N \sum_{i=1}^L N_i \gamma_{media\ (i)}$$

σ_{media}^2 = Estimador de la varianza poblacional.

$$\sigma_{media}^2 = 1 / N^2 \left[\sum_{i=1}^L N_i (N_i - n_i) / N_i (s_i^2 / n_i) \right]$$

w_i = fracción del tamaño de cada estrato a la población total.

$n_i = n w_i$ muestra del estrato i como fracción del tamaño de la

muestra total n . $i = 1, \dots, L$

Tamaño de la muestra para estimar la media poblacional μ con un límite B para el error de estimación

$$n = \frac{\sum_{i=1}^L N^2_i \sigma^2_i / w_i}{N^2 B^2 / 4 + \sum_{i=1}^L N_i \sigma_i^2}$$

Si se sabe el costo unitario de obtener una información en el estrato i, entonces la fórmula para encontrar el número de elementos de la muestra que minimice el costo total será:

c_i = costo unitario de una obtener una información en el estrato i

$$n = \frac{\left(\sum_{i=1}^L N^2_i \sigma^2_i / \sqrt{\sigma_i} \right) \left(\sum_{i=1}^L N_i \sigma_i \sqrt{c_i} \right)}{N^2 B^2 / 4 + \sum_{i=1}^L N_i \sigma_i^2}$$

6.4. Muestreo por conglomerados

Los conglomerados son subconjuntos de la población que tienen la propiedad de ser internamente heterogéneos y entre ellos lo más homogéneos posibles.

Una muestra por conglomerado es una muestra aleatoria en la cual cada unidad de muestras es una colección o conglomerado de elementos.

Este muestreo es un diseño efectivo para obtener una cantidad especificada de información al costo mínimo bajo las siguientes condiciones:

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

Elementos en un muestreo por conglomerados:

N = Número de conglomerados en la población.

n = Número de conglomerados seleccionados según Muestreo Aleatorio

m_i = Numero de elementos en el conglomerado i , $i = 1, \dots, N$

m_{media} = Tamaño promedio del conglomerado

$$m_{media} = 1 / n \sum m_i$$

$M = \sum m_i$ Número de elementos de la población.

$M_{media} = M / N$ tamaño promedio del conglomerado en la población.

γ_i = Total de todas las observaciones en el i -ésimo conglomerado.

El **estimador de la media poblacional** para el muestreo por conglomerado será:

$$\bar{\gamma} = \frac{\sum_{i=1}^n \gamma_i}{\sum_{i=1}^n m_i}$$

El **estimador de la varianza poblacional** en el muestreo por conglomerados será:

$$V = (N - n) / (N n m_{media}^2) \sum_{i=1}^n (\gamma_i - \gamma_{media} m_i)^2 / (n - 1)$$

6.5.- Muestreo sistemático

La muestra se elige mediante la selección de un punto de inicio aleatorio y después la elección de cada i -ésimo elemento en sucesión a partir del marco de la muestra.

Es similar a muestreo aleatorio simple en que cada elemento tiene una probabilidad de elección idéntica y conocida.

El muestreo sistemático es más susceptible de ser más preciso que el muestreo aleatorio simple.

Una muestra obtenida al seleccionar aleatoriamente un elemento de los primeros k elementos en el marco y después cada k -ésimo elemento se denomina muestra sistemática de 1 en k .

Por ejemplo: A partir de una lista de 100 establecimientos educacionales, deseamos seleccionar una muestra probabilística de 20 colegios. La forma de hacerlo sería:

Si dividimos 100 entre 20 obtendremos 5, que es el salto sistemático.

Se extrae un número al azar entre 1 y 5. Supongamos 2.

Se incluyen en la muestra los colegios numerados 2, 7, 12, 17, 22,.....97.

6.6.- Otros Muestreos:

Muestreo no probabilístico: Dependen del juicio personal del investigador, quien puede decidir de manera arbitraria o conciente que elementos va a incluir en la muestra.

Las muestras no probabilísticas se las denomina también muestras dirigidas, y suponen un procedimiento de selección informal y un poco arbitrario. Algunas de las técnicas más utilizadas en este tipo de muestreo son: Muestreo casual, intencional, por cuotas, por bola de nieve entre otros.

Muestreo casual: Se toma los elementos en forma casual. La selección de las unidades de muestreo se seleccionan porque se encuentran en el lugar adecuado en el momento oportuno.

Es la menos costosa y la que necesita menor tiempo. Los elementos se seleccionan de acuerdo al gusto del investigador.

Esta técnica se utiliza con frecuencia en proyectos de investigación de mercados comerciales.

Muestreo intencional: El investigador selecciona los elementos que a su juicio son representativos, lo cual exige del investigador un conocimiento previo de la población.

Muestreo por cuotas: El investigador selecciona los elementos representativos a su juicio. La muestra ha de ser proporcional a la población y en ella deberán tenerse en cuenta las diferentes categorías.

Es un muestreo en dos etapas: La primera etapa consiste en el desarrollo de categorías de control o cuotas de los elementos de la población (categorías relevantes como sexo, edad, raza, nivel socioeconómico, etc.), por lo general las cuotas se asignan de la misma manera en que está conformada la población.

En la segunda etapa, los elementos de la muestra se seleccionan de acuerdo a la conveniencia o juicio del investigador.

Luego de que se asignan las cuotas hay mucha libertad en la selección de los elementos de la muestra, el único requerimiento es que cumplan con las características de control.

Muestreo por método de bola de nieve: Se selecciona al azar un grupo de entrevistados por lo general en forma aleatoria. Después de la entrevista se les pide que elijan a otros que pertenezcan a la población meta de interés, es decir que los siguientes se eligen con base en las referencias o información que proporcionan los entrevistados iniciales.

La ventaja de este método de muestreo es que aumenta la probabilidad de encontrar la característica deseada en la población. Da como resultado una varianza de muestra relativamente baja y los costos son reducidos.

Errores en la Muestra.

Para que una muestra proporcione datos confiables, éstos deben ser representativos de la población, es decir, que los errores del muestreo deben ser relativamente pequeños para que éste no pierda validez. Ninguna muestra da garantía absoluta en relación con la población de donde ha sido extraída, de ahí, la importancia de poder determinar el posible margen de error y la frecuencia de los mismos dentro del conjunto de datos.

Generalmente se presentan dos tipos de errores: **sistemáticos o no muestrales y de muestreo.**

Errores muestrales: Cualquiera que sea el procedimiento utilizado y la perfección del método empleado, la muestra siempre diferirá de la población. A esta diferencia entre la población o universo y la muestra se le denomina error de muestreo.

El error de la muestra es la diferencia entre μ (que es el parámetro de la media para la población) y \bar{x} (las estimaciones de la media para la muestra).

Por el solo hecho de tomar una muestra y no la población ya existe un error muestral.

Errores no muestrales: Son aquellos que se presentan por causas ajenas a la muestra.

Situaciones inadecuadas: Se presentan cuando el investigador tiene dificultades para obtener la información y la sustituye por la que más fácilmente está a su alcance, que no siempre es la más confiable.

Insuficiencia en la recolección de datos: Hay distorsión por falta de respuestas, o respuestas inadecuadas, ya sea por ignorancia o falta de datos relativos a los elementos incluidos.

Errores de cobertura: Cuando no se han incluido elementos importantes y significativos para la investigación que se realiza.

