



DISEÑOS CUANTITATIVOS: ANÁLISIS E INTERPRETACIÓN DE LA INFORMACIÓN

Miguel Aigner A.

Abstract

The course, to write this educational material, is that the traditional teaching model used to teach research methodology integrates not specific to the process, actions and exercises to see in practice the use of tools such as statistics and computer science, that day by day, they have become key assistants of social research.

Resumen

El supuesto, al escribir este material docente, es que el modelo pedagógico tradicional utilizado para enseñar metodología de la investigación no integra al proceso concreto, acciones y ejercicios que permitan ver en forma práctica la utilización de herramientas como la estadística y la informática, que día a día, se han convertido en auxiliares importantes de la investigación social.



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

Miguel Aigner A.

Centro de Estudios de Opinión -CEO-

Facultad de Ciencias Sociales y Humanas

Universidad de Antioquia

Telefax 210 57 75

ceo@quimbaya.udea.edu.co

<http://ceo.udea.edu.co>

ISBN 655-318-3

Fondo Editorial CEO

Derechos reservados.

Prohibida la reproducción total o parcial de esta obra sin la autorización escrita del autor y del Centro de Estudios de Opinión de la Universidad de Antioquia.

.

Impreso y hecho en Colombia

ceo@catios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

Índice	3
Capítulo I: Introducción sobre los diseños cuantitativos	9
1. Introducción	10
2. Objetivos	10
3. Desarrollo	10
3.1. Diseño operativo de la investigación	11
3.1.1. El planteamiento de un problema de investigación	11
3.1.2. La formulación de hipótesis de investigación	14
3.1.3. El proceso de comprobación de hipótesis	15
3.1.4. Operacionalización de hipótesis	17
Ejercicio de autoevaluación N° 1	19
Capítulo 2: Tipos de diseños de investigación social	23
1 Introducción	24
2. Objetivos	24
3. Desarrollo	24
3.1. Tipo de diseños experimentales	24
3.2. Tipo de diseños no-experimentales	25
3.3. La calidad interna de la información	26
- Ejercicio de autoevaluación N° 2	29
Capítulo 3: El muestreo	30
1. Introducción	30
2. Objetivos	30
3. Desarrollo	31
3.1. Determinación de la muestra	31

3.2. Métodos de muestreo	31
3.2.1. Muestreo aleatorio o probabilístico	32
3.2.1.1. Muestreo aleatorio probabilístico simple	33
3.2.1.2. Muestreo sistemático	33
3.2.1.3. Muestreo estratificado	34
3.2.1.5. Muestreo estratificado por conglomerados	37
3.2.2. Muestreo no aleatorio – no probabilístico	40
3.2.2.2. Muestreo por cuotas	40
Capítulo 4: La recolección de la información	41
1. Introducción	41
2. Objetivos	41
3. Desarrollo	41
3.1. La medición en los diseños cuantitativos	41
3.1.1. Los niveles de medición y los estadísticos apropiados a cada nivel	42
3.1.2. Conceptos básicos de medición	48
3.1.2.1. Características de los niveles de medición	48
3.1.2.2. Criterios de medición	50
3.2. Instrumentos de recolección de información	54
3.2.1 Conceptos básicos de la recolección de información	55
3.2.2. Técnicas de recolección de la información	59
3.2.2.1. Medición escalar	59
3.2.2.2. La entrevista social	65
3.2.2.3. La encuesta social	65
-Ejercicio de autoevaluación N° 3	83
Capítulo 5: Sistematización de la información	84

1. Introducción	84
2. Objetivos	84
3. Desarrollo	85
3.1 Sistematización de la información	85
3.1.1. Plan de análisis de los datos	85
3.1.2. El proceso de codificación de la información obtenida	86
3.1.4. Generación de archivos de trabajo	91
3.1.4.1. Manejo operativo del SPSS	97
Capítulo 6: Análisis e interpretación de los datos cuantitativos	96
1. Introducción	96
2, Objetivos	96
3. Desarrollo	97
3.1. Análisis descriptivo de los datos	98
3.1.1. Análisis estadístico de los datos	102
3.1.1.1. Estadísticos para medir la localización de un conjunto de datos	104
3.1.1.2. Estadística para analizar la dispersión de un conjunto de datos	107
3.1.1.3. Selección de las técnicas estadísticas en el análisis de datos	109
3.1.2. Representaciones gráficas de las distribuciones de frecuencia	115
3.1.2.1. Tipos de gráficos	116
3.1.2.2. Limitaciones en el uso de figuras o gráficos	117
3.1.3. La interrelación entre variables	115
3.1.3.1. Tablas de contingencia y medidas de asociación	118
3.1.3.2. Análisis de datos tabulados	118
3.1.4. Inferencia estadística	132
3.1.4.1. Prueba de hipótesis	135
3.1.4.2 Selección del test estadístico	136
3.1.5. Medidas de relaciones entre variables	148



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

3.1.5.1. Medidas de asociación entre variables cualitativas	149
3.1.5.2. Asociación entre variables nominales	155
3.1.5.3. Asociación entre variables ordinales: coeficiente de Spearman	158
3.1.5.4. Análisis de regresión	165
3.1.5.5. Coeficientes de correlación	169
3.1.6. Análisis multivariado	173
3.1.6.1. Regresión múltiple	173
Bibliografía	175
Glosario de términos	178
Anexos	181

- INTRODUCCIÓN

Generalmente en el proceso concreto de construcción del conocimiento en el campo de la investigación social se habla de diseños cuantitativos y cualitativos: El diseño cuantitativo es una forma de aproximación sistemática al estudio de la realidad que se apoya principalmente en categorías numéricas y utiliza preferentemente como base una información factible de cuantificar realizando el análisis a través de diferentes formas de interrelacionarlas. El énfasis en el análisis de los datos cuantificados se basa en las tendencias que muestra el comportamiento de ellos.

El supuesto, al escribir este material docente, es que el modelo pedagógico tradicional utilizado para enseñar metodología de la investigación no integra al proceso concreto, acciones y ejercicios que permitan ver en forma práctica la utilización de herramientas como la estadística y la informática, que día a día, se han convertido en auxiliares importantes de la investigación social.

Este Módulo está formada por seis capítulos, en las que se hace énfasis, en el proceso investigativo que prioriza el uso de una información cuantificada, desde el planteamiento del problema hasta el análisis de la información.

El primer Capítulo es una introducción a los diseños cuantitativos y aporta una información general sobre los diseños de investigación que se utilizan preferentemente, mientras que en el Capítulo dos se hace una presentación de los diferentes tipos de diseños de investigación utilizados en la recolección y medición de información cuantitativa. El Capítulo tres presenta una introducción a los diseños muestrales. El Capítulo cuatro muestra algunos conceptos y técnicas básicas en la recolección de datos haciendo énfasis en la técnica de la encuesta social. El Capítulo cinco, da cuenta del proceso de sistematización, crítica y validación de la información, codificación, digitación

y procesamiento de la información obtenida. El Capítulo seis trata del análisis de datos cuantitativos. Aquí se sintetiza lo que se debe saber para poder seleccionar la prueba estadística apropiada para cada tipo y problema de investigación. También, se enfatiza el manejo de la información para su uso en procesos de sistematización utilizando las herramientas que nos facilita el programa SPSS¹ y se revisan algunas herramientas estadísticas asociadas al análisis descriptivo, inferencial y asociativo.

Cada Capítulo esta organizado en unos contenidos que van acompañados de *ejercicios resueltos y de autoevaluación* que permiten practicar los contenidos teóricos y metodológicos y asegurarse de haberlos comprendido.

Es igualmente importante seguir las "*Notas metodológicas*" que son sugerencias procedimentales, algunas definiciones operacionales u observaciones metodológicas generales.

En la sección titulada *Procedimientos SPSS se detalla paso a paso como se utilizan las herramientas que ofrece este programa de manejo y gestión de datos cuantificados.*

- OBJETIVOS

Se pretende responder a la tendencia vigente en torno al énfasis profesional y a la especialización – conducente a mejorar la eficacia y eficiencia del desempeño profesional. Presentar el amplio panorama de los diseños cuantitativos como una variedad de estrategias adoptables, en forma creativa, imaginativa y sagaz, a los distintos objetos de investigación, de tal forma, que los llamados *problemas teóricos,*

¹ SPSS es un programa de archivo y gestión de datos numéricos que integra procedimientos estadísticos con gráficos, asistiendo al usuario en la manipulación, organización y archivo mediante menús y submenús.

metodológicos, técnicos y empíricos, que suelen presentarse en todo proceso investigativo, no obstaculicen su aplicación.

Familiarizar a los estudiantes con las herramientas de la estadística y la informática aplicados a la investigación social para que puedan utilizar estas herramientas, en el análisis de los datos más representativos, entender la lógica de la utilización de ciertos estadísticos y ser capaces de *leer* e interpretar los resultados obtenidos.

- CONCLUSION

Como se afirma en los párrafos anteriores, se trata de dar una visión técnica y práctica del proceso de investigación social que se apoya preferentemente en los diseños cuantitativos tratando de acercarnos a la práctica real del investigador. También se insiste en la tendencia de combinar diversos diferentes diseños metodológicos y la necesaria complementariedad entre teorías, metodologías y técnicas dada la complejidad del estudio de lo social. Igualmente, se reitera la necesidad de familiarizar al investigador social en una estrategia investigativa que no solo evite las dificultades metodológicas de los diseños cuantitativos, sino también la unilateralidad del momento cuantitativo.

En el módulo se aborda la presentación de las herramientas estadísticas más útiles en el tratamiento cuantitativo de las variables de una investigación, buscando dejar claros una serie de conceptos y técnicas.

CAPÍTULO 1. INTRODUCCIÓN SOBRE LOS DISEÑOS CUANTITATIVOS

1. Introducción

El punto de partida de la investigación social, es desde luego, la existencia de un problema que requiere alguna forma de respuesta, de explicación o de solución. El propósito final de este proceso es justamente lograr una respuesta lógica tanto desde el punto de vista de la formalidad teórica como metodológica, con esto el hecho problemático es incorporado o puesto en relación de implicación con la teoría pertinente.

La aplicación de métodos y técnicas específicas en la investigación social no se hace de manera rígida o mecánica de tal modo que la creación de conocimientos resulte de la sumisión estricta a un conjunto de reglas de procesamiento preestablecidas e inmodificables. El término *método científico* es entendido aquí como un concepto general que comprende muchas y distintas maneras de abordar un problema, de conocer, de acercarse a la realidad, guiadas por principios o modelos teórico-metodológicos que permiten en definitiva considerar los resultados logrados como conocimiento científico.

Esta forma de ver las estrategias investigativas del diseño cuantitativo resalta como idea, que la investigación social, es un proceso donde la creatividad, el compromiso del analista, la imaginación, las ideas, la argumentación, la perspicacia y la sagacidad son instrumentos válidos y confiables. Toda *formalización metodológica* tiene sólo un carácter instrumental, indicativo, protocolario, provisional, que facilite el desarrollo del proyecto de investigación.

El investigador debe estar atento, de manera continua, a los posibles errores y desviaciones a las cuales puede conducirlo el uso indiscriminado de algunas técnicas de investigación. En este caso, una sólida formación metodológica le permitirá una selección

adecuada de algunos de estos procedimientos o la limitación de su aplicación a ciertos temas y aspectos de la realidad social.

La metodología que se sigue para realizar un trabajo investigativo, es algo que se necesita conocer y aprender. Requiere de cierto conocimiento de la realidad en estudio; cierto *capital de base* que le permita al analista plantearse, al menos inquietudes y posibles respuestas sobre el tema, porque no existe un conjunto formalizado "único" de "protocolos-normas" que se pueden aplicar mecánicamente a cualquier situación problemática.

La observación anterior- junto a otras que se hacen luego, son de especial relevancia para la utilización de este módulo. En efecto, en los capítulos siguientes se expondrán diversas técnicas de recolección y análisis de datos apropiados para ciertos niveles de realidad social y destinada a tratar propiedades susceptibles de expresión cuantitativa.

2. Objetivos

Actualizar conocimientos y competencias en los campos de la metodología y técnicas de investigación social. Analizar casos que ejemplifiquen la utilización de métodos y técnicas de la investigación social.

3. Desarrollo

Considerar la metodología como una simple lista de pasos o etapas estandarizadas sería ir en contra de la esencia misma de la metodología científica, porque ésta no es un conjunto de recomendaciones que se puedan aplicar siempre. Sin embargo, existen elementos específicos comunes a toda investigación (especialmente en los llamados diseños cuantitativos) que ayudan a pensar y planear mejor, tales como *los aspectos*

científicos-técnicos de un proyecto - Qué hacer y como hacerlo - y los aspectos de administración y control del proyecto - Estrategia operativa.

Reconocida la diferencia entre el accionar real del investigador y la exposición didáctica de la misma, es posible distinguir etapas en el proceso de investigación. Estas etapas son las siguientes:

3.1. Diseño operativo de la investigación

En todo proceso investigativo hay que definir y diseñar un plan que sirva de guía para el desarrollo del estudio. El diseño de investigación implica no sólo parte del planteamiento del problema sino también la parte operativa que tiene que ver con la especificación de la metodología y técnicas para recolectar y validar la información necesaria para darles una respuesta.

El plan operativo de la investigación comprende un conjunto de actividades que parten de:

- Identificación, análisis y formulación del problema que se va a investigar.
- Determinación del tipo de diseño de investigación que se va a utilizar.
- Especificación de la hipótesis del estudio (sí las hay explícitas).
- Diseño de *la matriz de variables – dimensiones*: definición, clasificación y operacionalización de las variables del estudio.
- Selección de las fuentes de información y, en su caso, diseño de la muestra.
- Diseño del cuestionario.
- Recolección y sistematización de los datos.
- Análisis e interpretación de los resultados.

3.1.1. El planteamiento de un problema de investigación

Se dice que un problema de investigación es una situación a la cual se le quiere encontrar respuesta o solución. En este momento de la investigación es posible que el analista social se vea enfrentado a dos cuestiones. Generalmente, en el discurso pedagógico, se insinúa que el proceso de investigación sigue un camino unidireccional - punto de partida y punto de llegada- pero es sólo parte del discurso, en la realidad uno va

de lo abstracto a lo concreto, o de adelante para atrás y viceversa; es el mismo proceso el que le va sugiriendo los pasos a seguir, no el manual.

¿Cuál es el problema y cuál es la finalidad última de la investigación? En la realidad esas diferentes concepciones-suposiciones acerca de cuál es el problema, cuál es su tratamiento teórico metodológico, cuál es la respuesta son una expresión de diferentes puntos de vista teóricos, metodológicos e ideológicos que se evidencian en los llamados paradigmas de investigación social.

Pero normalmente este hecho o proceso social que llama nuestra atención, emerge de un tema global mediante un proceso deductivo. Una primera aproximación nos llevará a *identificar* y aislar nuestra situación problemática con el fin de especificar que se va concretamente a investigar.

Supongamos que el director de bienestar estudiantil en una universidad pública está interesado en conocer el grado de aceptación de los programas culturales por parte de la comunidad estudiantil. Este es un problema de investigación identificado, lo que implica que está delimitado dentro de una temática global y se tiene conciencia de la presencia de un vacío específico de conocimiento.

Una vez identificado el problema, se le somete a un *proceso de análisis*; aquí lo que hace el investigador es plantearse una serie de interrogantes, tales como los siguientes:

"En cuáles circunstancias aparece el problema; qué elementos básicos lo componen; [...] qué interrelaciones existen entre los elementos componentes: cuáles elementos dependen de otros; qué mediaciones existen en esas relaciones; qué aspectos del problema se conocen; cuáles están identificados;

cuáles faltan por hacerlo; qué tipo de explicaciones o modelos conceptuales nos permiten tener una mejor comprensión de ellos... es decir, se trata de desmenuzar el problema” (Facundo, Ángel, 1987:22)

Contextualizado el problema de investigación, delimitados y examinados sus elementos y/o variables y definida la forma como se interrelacionan; el paso siguiente es *su formulación*, es decir, se debe precisar que, es lo que se va estudiar, tanto los elementos, características o variables que en él intervienen, como las relaciones existentes entre ellas y que están a la espera de una explicación cierta y comprobables.

La formulación del problema de investigación es una actividad fundamentalmente práctica (se aprende haciéndolo) ya que el proceso arranca desde la identificación de un problema de investigación, cuestión que requiere no sólo de una capacidad de *ver* los problemas y de *una actitud* de respuesta o solución, sino que además hay que sumarle una cierta capacidad de imaginación, de sentido común y manejo creativo de esa parte de la realidad que se quiere escudriñar.

En el planteamiento de un problema se concretizan básicamente dos situaciones. Por un lado, unas ideas que comprenden un esquema o modelo de qué hay que investigar (objeto general de estudio); del cómo hacerlo y la forma de *interpretar* los hechos (si bien en muchas ocasiones tales *esquemas o modelos* no se explicitan) y, por otro, *una metodología* -procedimientos, protocolos, técnicas- que nos van a permitir encontrar respuesta.

Al investigador se le pueden presentar diversas alternativas teórico-metodológicas dentro de las cuales es posible encontrar *la ansiada respuesta*. Surgirán entonces varias

preguntas a responder, por ejemplo en el problema planteado anteriormente acerca del grado de aceptación de los programas culturales por parte de la comunidad estudiantil:

¿Qué tipo de información requerimos para responder a lo planteado?

¿A qué tipo de programas culturales nos estamos refiriendo?

¿Qué entendemos por grado de aceptación?

¿Es un estudio comparativo? Por ejemplo con el grado de aceptación de años anteriores.

¿Es un estudio evaluativo?

Cada una de estas preguntas requiere de diferentes diseños metodológicos y distinto tipo de información. Igualmente antes de recolectar la información se deberá diseñar, de acuerdo a la técnica seleccionada, una guía de entrevista, una de observación o un cuestionario. También habrá que planificar el proceso de recolección y análisis de la información y por último, analizar e interpretar los resultados, a la luz de un marco teórico.

3.1.2. La formulación de las hipótesis de investigación

Toda investigación se orienta por hipótesis que hacen el papel de guías de trabajo, las cuales pueden ser explícitas o implícitas.

Las principales funciones de una hipótesis son:

- Conducir la investigación: *"La hipótesis [...] guía toda su marcha, desde la caracterización de la población a la cual se refiere y necesita para su contratación, la información a recoger, el manejo del trabajo de campo y las técnicas adecuadas para el análisis de los datos."* (Briones, 1988:54)
- Especificar cuáles hechos se observan y que relación establecerá entre ellos.
- Ofrecer una suposición o explicación tentativa al problema de investigación.

Supongamos que se ha identificado un problema de investigación como el que más adelante se presenta; donde es posible apreciar cómo en el mismo planteamiento del problema se enuncian explícitamente algunas respuestas, se hacen algunas afirmaciones o suposiciones tentativas o mejor dicho, una hipótesis que necesitamos comprobar a través de un proceso investigativo.

- Ejemplificación

"Es conveniente desmontar el sentido de pertenencia y arraigo sobre aspectos físicos, y resaltar más bien la necesidad de sondear que, elementos son posibles de incorporar para restablecer o crear formas de convivencia ciudadana, con el fin de lograr una ciudad mejor para vivir. Bajo esta perspectiva, los procesos demográficos, los cambios en la organización familiar, en el sistema educativo [...] en la cotidianidad, configuran el universo de factores que en la ciudad de Medellín asumen perfiles diferentes, no sólo por determinaciones socioeconómicas y socio espaciales, sino ante todo por la perceptible incoherencia de los patrones y valores de comportamiento y cultural que encarna cada generación según el modelo y la temporalidad inscrita." (Ruiz, Jaime, et. al. 1991:24-25)

Alrededor de este planteamiento surgen una serie de preguntas o interrogantes que presentados en forma de hipótesis se pueden probar o refutar, por ejemplo, a partir de un estudio de opinión.

Como se aprecia antes de formular una hipótesis es necesario reunir suficientes antecedentes y conclusiones teóricas, comparar un número tan grande como sea posible de información asociada disponiendo para ello de la documentación pertinente. Es decir, para establecer una hipótesis se requiere seguir el camino de lo conocido a lo supuesto.

3.1.3. El proceso de comprobación de hipótesis de investigación

Una hipótesis comprobable debe reunir las siguientes características:

- Ser conceptualmente operacionalizable, es decir, debe tener un referente empírico observable.
- Permitir la utilización de técnicas que faciliten las observaciones.
- Estar relacionada con un marco de referencia teórico-conceptual que facilite el análisis e interpretación de los datos que van a permitir comprobar o refutar la hipótesis².

Volvamos a la ejemplificación anterior, acerca del Programa de seguridad comunitaria, en cada una de las afirmaciones reseñadas en los párrafos anteriores, hay una respuesta tentativa, una afirmación, una suposición acerca de la situación planteada, pero además hay una noción implícita de un factor causal y de una consecuencia: los conceptos de variable dependiente y de variable independiente reflejan esta situación³.

- Ejemplificación

Nuestra hipótesis es que la violencia responde a una doble ruptura: la primera es aquella entre la clase política, en cuyo seno buscan replegarse los dirigentes comunitarios reivindicativos y la vida comunitaria que cotidianamente rechaza a los políticos tradicionales; la segunda es la ruptura entre *los barrios populares*, suertes de gethos adonde están confinados los pobres urbanos, y el resto de la sociedad, cuya cohesión

² En los diseños cuantitativos, generalmente, cuando se habla de comprobar hipótesis significa someterla a prueba por medio de observaciones.

³ Por *variable* entenderemos aquellos enunciados formados por significados teóricos que nos permiten transmitir ideas referidas a características de sujetos o hechos que de alguna forma varían o cambian.

parece concertarse cada vez más sobre la base de esta discriminación geográfica y social.

En este caso hipotético, los factores causales o determinantes -variable independiente - serían la llamada *ruptura entre la clase política y la vida comunitaria y la segregación geográfica y social*. La consecuencia sería un antimovimiento social, cuya expresión más patente es la violencia barrial, especialmente entre los jóvenes.

Variable independiente⁴	Variable dependiente⁵
Rupturas políticas Segregación socio – política	Tipo y grado de violencia barrial

El analista social debe demostrar si la relación enunciada en la hipótesis es válida o inválida. En síntesis, cualquiera sea la forma del enunciado de una hipótesis, su intención debe ser la de comprobar o refutar su afirmación o respuesta tentativa. En la investigación cuantitativa, las hipótesis no siempre son enunciadas explícitamente, aún así, es necesario que la pregunta central del problema de investigación este presente, así sea sólo en la mente del analista, con el fin de no perder la orientación de lo que se quiere.

⁴ Por variable independiente entenderemos esa propiedad, esa situación o hecho que se supone es el causante o explica el efecto señalado.

⁵ Por variable dependiente entenderemos los efectos o la consecuencia de un hecho o proceso.

3.1.4. Operacionalización de hipótesis de investigación

Un problema común en un enunciado hipotético es la existencia de conceptos que pueden ser entendidos de diversas formas por los interlocutores, lo cual va a determinar graves sesgos en la información recolectada. Hay otros conceptos que por su abstracción son de difícil operacionalización tales como por ejemplo; *autoridad, tranquilidad, felicidad, ansiedad, eficiencia, miedo, etc.*

Con el fin de probar una hipótesis, se hace necesario, no sólo tener una visión integral del problema a investigar, sino también el marco de referencia teórico conceptual que lo define. En nuestro ejercicio, acerca de la relación entre rupturas políticas, segregación socio-espacial y violencia barrial (medida por índices de conductas delincuenciales), implícitamente hay una suposición en el sentido de que la reconstrucción del tejido social y político permitirían erradicar a los antisociales y sus manifestaciones delincuenciales; esta afirmación parece ser muy explícita pero en la práctica no es un planteamiento operacional: Habría que definir **operacionalmente** los conceptos como *ruptura política, segregación socio-espacial, erradicación, conductas delincuenciales*, de tal forma que permitan entender claramente que se entiende por ellos y cómo van a ser **traducidos** en referentes empíricos observables en el proceso de diseño del cuestionario.

- Definición, tipos y medida de las variables

Una variable es una característica o propiedad de un sujeto, hecho social o situación cuyos valores que son objeto de estudio pueden variar y mostrar diferentes valores de una unidad de análisis a otra.

Las variables pueden asumir diferentes *valores* lo que se puede observar en la siguiente ejemplificación:

Variables

Valores

Nivel o grado de escolaridad

Primaria incompleta
Primaria completa
Secundaria incompleta
Secundaria completa
Educ. Superior incompleta
Educ. Superior completa

Tipo de variables

Según *la relación causa - efecto* analizada, las variables pueden dividirse, de acuerdo con el papel que desempeñen en la relación causal, en independientes, dependientes e intervinientes:

- **Variable independiente.** Es una variable que se utiliza para explicar el comportamiento de otra variable dependiente. Se denomina también variable explicativa o predictora. Por ejemplo, el nivel de ingresos puede ser explicado por el nivel de escolaridad, que en este caso es la variable independiente.

- **Variable dependiente.** Es aquella cuyo comportamiento es explicado por otra u otras independientes. Se la denominaba también variable a explicar, por ejemplo, el comportamiento lector (variable dependiente) puede ser explicado por la edad del individuo.

- **Variable interviniente.** Supone que la relación entre dos variables está mediada por otra u otras que determinan o influyen, si es el caso, los efectos o las características de las primeras. Cabe preguntarse desde ya, acerca de como decidir: *¿Cuál es la variable interviniente?* Ya que es evidente que en cada hipótesis hay potencialmente muchas variables que pueden intervenir. La respuesta está en la pregunta acerca de cuáles factores (de acuerdo al criterio teórico-metodológico del analista) están más relacionados, tanto con la variable Independiente como con la variable dependiente.

En función de *los valores cuantitativos que pueden adoptar*, las variables se clasifican en los siguientes tipos:

- **Variable continua.** Es aquella que puede tomar cualquier valor numérico. Por ejemplo, el monto de los salarios de una población en estudio, rendimiento académico, expresado en promedios.

- **Variable discreta.** Es aquella que sólo puede tomar un número finito de valores distintos. Por ejemplo, el número de hijos.

- **Variable dicotómica.** Aquella que sólo puede tomar dos valores, por ejemplo, el sexo (hombre - mujer), una afirmación o negación (sí - no), la evaluación de un test (aprobado – reprobado) o una apreciación o creencia (verdadero - falso).

- Ejercicio de autoevaluación N° 1

En un estudio sobre *los programas de seguridad barrial* se hace referencia a las políticas y las circunstancias socio-económicas más determinantes en la contención de la violencia barrial. Es posible suponer que hay otros factores socioeconómicos asociados al proceso

que estamos analizando y que pueden tener un impacto independiente de la existencia o no de programas de seguridad comunitaria, tal como el desempleo en grupos de jóvenes. Esta clase de factores intervinientes deben ser tenidos en cuenta cuando se planea el diseño de la investigación, aún cuando se decida que ellos no son lo suficientemente importantes para ser tenidos en cuenta en el análisis. Igualmente, factores tales como los culturales e ideológicos, que pueden ser un reflejo de opiniones, actitudes y conductas formadas por los medios de comunicación, deben ser tenidos en cuenta en los resultados esperados.

De la siguiente hipótesis del estudio reseñado que hace referencia a cómo las formas no-tradicionales de solidaridad barrial han aminorado ciertas conductas delincuenciales, señale cual es la variable independiente y la variable dependiente.

Variable independiente _____

Variable dependiente _____

Mencione tres variables intervinientes o de control que permitirían *poner a prueba* la consistencia de la relación original enunciada en la primera hipótesis

Variable interviniente 1. _____

Variable interviniente 2. _____

Variable interviniente 3. _____

- Unidades de análisis

En los diseños cuantitativos la unidad de análisis es un dato -observable y externo- que se convierte en la unidad de información básica. Es un concepto referido al objeto de estudio, por ejemplo, individuos, grupos, regiones, hechos, procesos. Su definición es un paso importante en el planteamiento del problema de investigación. Al respecto se recomienda tener la mente abierta en cuanto a la naturaleza de las unidades de análisis; no siempre en la investigación social las unidades están definidas en términos de individuos o conjuntos de individuos. *Las unidades* pueden definirse en términos de los productos de los individuos, sus comportamientos o pueden ser de entes geográficos o físicos. Son pues unidades de análisis con carácter de artefactos o productos del hombre, los libros, periódicos, automóviles, viviendas entre muchos otros; así como países, regiones, ciudades, barrios, etc. Cada una de esas unidades supone la existencia de una población o universo de tales objetos y cada unidad puede ser descrita, clasificada o caracterizada en términos de las dimensiones significativas para el estudio.

Hay que mencionar también la posibilidad de utilizar otras unidades no naturales sino construidas, como son las categorías económicas, administrativas, etc., constituidas por elementos sin relación de interacción o sin estructura, como grupos de edad, de espacio o de tiempo (meses, días, años). En ocasiones es posible -y necesario- manejar varias unidades de análisis en una investigación, pero es preciso definir de antemano cuales conclusiones deben derivarse de una u otra.

- Ejemplificación

Problema de investigación: *La deserción académica estudiantil en los primeros semestres en la Universidad de Antioquia.*

Antecedentes

Uno de los mayores problemas que afecta la educación superior en Colombia, es la deserción académica de los estudiantes universitarios en sus primeros semestres de estudios. Entre las posibles causas de la llamada mortalidad académica están: La falta de orientación profesional, el bajo nivel de la educación secundaria, los bajos recursos económicos de las familias o el bajo rendimiento académico en los primeros semestres universitarios. Muchos de estos estudiantes que desertan de la Universidad se vinculan al mercado laboral en empleos mal remunerados, o se inscriben en institutos tecnológicos en donde hacen una carrera tecnológica, a la cual debieron haber llegado directamente, sin la experiencia frustrante de la Universidad.

Unidades de análisis, variables y valores

1. La unidad de análisis (sujeto-objeto) del problema son *los estudiantes universitarios en sus primeros semestres académicos de estudio.*
2. La mortalidad académica estudiantil es un problema que está determinado por los siguientes factores que la determinan (variables independientes): *la falta de orientación profesional, la baja calidad de la enseñanza secundaria, los bajos recursos económicos de las familias y el bajo rendimiento académico en los primeros semestres universitarios.*
3. Las familias de los estudiantes que ingresan a la Universidad, podrían dividirse en familias de recursos económicos: *bajos, medio-bajos; medios; medio-altos; altos.*

CAPÍTULO 2. TIPOS DE DISEÑOS DE INVESTIGACIÓN SOCIAL

1. Introducción

En términos metodológicos se habla de diseños experimentales y no experimentales. Hay fundamentalmente dos elementos que diferencian los diseños de tipo experimental o cuasi-experimental de los denominados *no-experimentales*.

La elección de un tipo de diseño u otro, depende del problema a resolver, de los objetivos y del contexto que rodea del trabajo de investigación. No sólo pueden combinarse sino también complementarse.

Generalmente, existen dos clases de factores que inciden sobre la validez de los resultados obtenidos con estos tipos de diseños de investigación: Los históricos, los cuales hacen referencia a sucesos o procesos acaecidos durante el tiempo del experimento y a la vez pueden haber tenido cierto grado de influencia en ese último; el efecto del factor instrumental⁶, el cual hay que controlar en los estudios de series de tiempo, en el sentido que la dimensión conceptual al ser operacionalizada en un proceso de medición realmente mida, en todas las circunstancias, lo que conceptualmente se quiere medir.

Finalmente, hay que dejar en claro, que estos criterios clasificatorios de los diferentes tipos de diseños, sólo tiene una justificación pedagógica. La práctica real de la investigación muestra generalmente inteligentes combinaciones de diseños que responden a los objetivos e intereses del investigador y no a la sujeción a protocolos metodológicos absolutos

⁶ Definición conceptual operacional.

2. Objetivos

Familiarizar al estudiante con los algunos de los tipos de diseños de investigación social.

3. Desarrollo

3.1. Tipos de diseños experimentales

Generalmente, estos diseños se utilizan para estudiar conductas sociales buscando controlar el impacto de factores externos en una relación hipotética de dos grupos seleccionados al azar, uno experimental y otro de control, midiendo el impacto del proceso u hecho en espacios temporales diferentes. En estos, hay un control más riguroso, se analizan *relaciones experimentales o controladas* y se pueden *manipular* las variables por separado o conjuntamente con otras para conocer sus efectos.

3.1.2. Diseño con un grupo de control no equivalente

En el cual los grupos experimentales y de control no han sido formados al azar; generalmente son grupos naturales, como por ejemplo, los funcionarios de la oficina de planeación y catastro municipal. El control de variables extrañas al experimento se hace mediante la elección de un grupo que tenga el mayor grado posible de características similares con el grupo experimental.

3.1.3. Diseño de antes y después con grupo de control

Con ellos se busca seleccionar -intencionalmente a los grupos experimentales y de control, con el criterio que reúnan en términos generales las características del grupo experimental.

3.1.4 Diseño de antes y después

Este tipo de diseño señala la necesidad de comparar opiniones teniendo como referente una información previa, ya que es evidente que un diseño descriptivo será incapaz de evaluar, por ejemplo, el grado de aceptación de la nueva reglamentación, sin tener un referente.

3.1.5. Post-test con solo comparación

En este diseño los sujetos del estudio son comparados con un grupo de control solamente después de haber participado en un programa, tal como el ejemplificado anteriormente. Obviamente no hay *Pre-test*. Los estudios de opinión son un típico ejemplo del diseño de *Post-test*.

3.2. Diseños no-experimentales

En la práctica investigativa de lo social es difícil lograr un control de todos los factores internos y externos que pueden invalidar una información. En aquellas situaciones en la que no nos es posible el control y manipulación de las variables es posible encontrar otros diseños metodológicos alternativos, como los diseños no-experimentales. Entre estos podemos mencionar:

3.2.1. Diseños descriptivos⁷: Tienen como objetivo describir características de una población sujeto de estudio y analizar su incidencia o interrelación en un momento dado. Son una fotografía de un individuo o colectivo.

3.2.2. Diseños correlacionales: Tienen como objetivo describir las relaciones de unos factores o variables en relación con otros, que son considerados como determinantes; es decir predice la relación entre dos variables, ninguna de las cuales es controlada o manipulada por el investigador. Se caracteriza porque permite identificar, medir,

⁷ Conocidos también como diseños transeccionales

interrelacionar múltiples variables simultáneamente, no importando establecer relaciones de causa-efecto entre ellas.

3.2.3. Diseños longitudinales: En ciertas circunstancias, el interés del analista es analizar cambios a través del tiempo de ciertas variables o en las relaciones entre éstas. Los diseños longitudinales recolectan información a través del tiempo en períodos especificados para hacer inferencias respecto “al cambio, sus determinaciones y consecuencias” (Hernández, Roberto (1991), p. 196). Un buen ejemplo de estos diseños son los sondeos electorales tipo panel que siguen las preferencias electorales a través del tiempo. Se habla de tres tipos de diseños longitudinales: De tendencia, de evolución de grupo y de panel. Se diferencian por las características de las unidades de análisis – grupos específicos o población en general.

Con los últimos, los de panel, también llamados *series de tiempos* en los cuales no se requiere de un grupo de control, consisten en una serie de mediciones periódicas que se hacen a las unidades de análisis del estudio, antes y después que se aplica el tratamiento. En este tipo de diseños se puede intentar mediatizar la influencia de los factores externos a través un sistema de medidas a lo largo del estudio.

3.3. La calidad interna de la información: La validez y confiabilidad de los datos

*La validez y confiabilidad*⁸ de un diseño cuantitativo se fundamenta básicamente en la posibilidad de controlar o tener en cuenta los siguientes factores:

⁸ Por confiabilidad se entiende la consistencia de los puntajes o la ausencia de errores en mediciones sucesivas. Por validez interna entenderemos aquí, el grado en que las mediciones empíricas reflejan la dimensión conceptual que se quiere medir.

- *Factor maduración.* Un diseño investigativo, cuya unidad de análisis, que como en el caso de la ejemplificación son escolares que mental y biológicamente van desarrollándose, deberá tomar en cuenta, que independiente de cualquier tipo de programa didáctico, los escolares van incrementando sus habilidades de todo tipo, situación que de no tenerse en cuenta determinará un sesgo en la medición.

- *Selección.* Así el grupo de estudio sea homogéneo en edad y nivel de escolaridad, no lo es en términos individuales, en el sentido de los procesos diferenciados de socialización familiar y profesional. Es dado pensar que hijos de profesionales asociados a áreas del conocimiento en las ciencias exactas tendrán habilidades mayores o más desarrolladas. Esta particular diferenciación sesgará los puntajes independientemente del programa didáctico.

- *Mortalidad.* Es de suponer que en un grupo como el del estudio señalado anteriormente, habrá mortalidad y deserción académica lo cual sesgará los resultados en la prueba del **post-test**.

- *La repetición del cuestionario.* Es posible que los sujetos del estudio, logren cierta habilidad en responder los cuestionarios, dada la familiarización que lograron al serle aplicado los *pre-test*, lo cual incidirá en los resultados del **post-test**. Es posible que los puntajes de medición tiendan a subir como producto de ciertas características del cuestionario del Post-test que faciliten las respuestas o porque la familiarización de los participantes del experimento con los asistentes de la investigación cree, una situación que influya en las respuestas acertadas.

- *El factor histórico.* También es posible que durante el periodo de tiempo que ha transcurrido en el estudio, influencias externas a la escuela, incentiven el desarrollo de

"habilidades matemáticas" en los escolares, por ejemplo, que en un canal cultural de televisión presenten programas dedicados justamente a desarrollar esas habilidades que se están midiendo con el experimento.

Ejemplificación

La Secretaría de Educación Municipal está interesada en mejorar el rendimiento académico de los estudiantes de los establecimientos educativos de enseñanza secundaria mediante el uso de programas pedagógicos innovadores. Se han propuesto por medio de un estudio evaluar un programa académico asistido por técnicas informáticas. Y se ha llegado a la conclusión que el desarrollo de la habilidad matemática de los estudiantes asistidos mediante programas tutoriales informáticos es mayor que la de aquellos que no han tenido acceso a ellos.

Se asume, para los efectos de esta ejemplificación, que el programa tutorial es didáctico, se ha definido operacionalmente el concepto de *habilidad matemática* y que se cuenta con un cuestionario que logra *medir* esa habilidad.

Desde el punto de vista metodológico del diseño de la investigación, se toma la primera decisión de poner a prueba la habilidad matemática del grupo de estudio, antes y después de haberlos sometido al experimento pedagógico, con el fin de intentar medir sus desarrollos en *habilidades matemáticas*.

Al final del Programa, ambos grupos serán sometidos a *una prueba post-test*; Es de suponer que el grupo experimental ganará en puntajes comparado con los obtenidos en *el pre-test* y que la ganancia-aumento del grupo de control tendrá un incremento menor.

Lamentablemente esta sencilla estrategia investigativa, seguramente nos llevará a que los resultados que encontremos no tengan mayor validez. Y esto puede explicarse básicamente por lo que los metodólogos denominan la no-confiabilidad en la calidad interna de la información⁹.

- Criterios metodológicos de validación de la información

Para superar estos sesgos y distorsiones se recomienda no solo tener en cuenta los factores que se enunciaron anteriormente, sino además, asegurar algunos criterios metodológicos de control, como los que se ejemplifican a continuación:

- Definir un listado de posibles participantes en el experimento, teniendo como base características socio-demográficas homogéneas, tales como, edad, lugar de residencia, estrato social, etc.
- Mediante una técnica de muestreo se selecciona un grupo de estudio; se divide el grupo: la mitad integrarán el grupo experimental y la otra mitad el grupo de control. A cada uno de los participantes de cada grupo se le aplicará un *Pre-test* para medir en ese momento sus habilidades matemáticas. Es de suponer que los resultados serán relativamente similares en cada grupo.

Teniendo presente estos factores internos y controlándolos se lograrán resultados más validos; pero el problema no se reduce sólo a estos factores sino que también tienen peso los llamados *factores de calidad externa*¹⁰, que se refieren a la dificultad de generalizar los resultados producto de una situación experimental controlada a otro segmento poblacional, sin tomar algunos criterios de control.

⁹ La calidad interna de la información depende de la confiabilidad y de la validez.

¹⁰ La calidad externa de la información depende de la representatividad, la cual tiene que ver con el grado en que los resultados, a partir de la muestra, pueden ser atribuidos a la población en general.

- Ejercicio de autoevaluación N° 2

1. ¿Cuáles son los factores más comunes que afectan *la calidad interna de la información* en estudios que utilizan diseños metodológicos de tipo: Post-test, Post-test con comparación y él de antes y después?

2. Supongamos que usted es un analista social que ha sido contratado para evaluar un programa de reinserción de jóvenes delincuentes, del cual, uno de los objetivos es dar una oportunidad a jóvenes no reincidentes para ser admitidos deben ser sometidos a un escrutinio de sus antecedentes y predisposición a participar en el Programa. Una vez admitidos, los cargos penales son cesados mediante un convenio entre las partes que compromete al participante a restituir mediante servicios comunitarios, los servicios recibidos. Partidarios del Programa arguyen que es menos costoso, para el erario público, la implementación de este tipo de metodologías de rehabilitación que seguir los trámites judiciales tradicionales; al igual, que es exitoso en términos de la no-reincidencia. Ante la posibilidad de que el programa se termine (situación no extraña a nuestras administraciones públicas) se le solicita al analista que reúna toda la información que pueda, apoyada en datos concretos representativos, confiables y validos que permitan mostrar que el programa sí está produciendo los resultados esperados.

Con base en el anterior problema de investigación construya *un diseño operacional*¹¹ provisorio que provea la información pertinente y metodológicamente adecuada para desarrollar el estudio. Asegúrese de ser exhaustivo acerca de las dimensiones - variables (preguntas básicas que Ud. deberá hacer) con relación al problema planteado y la relación del diseño de la investigación con esas dimensiones preguntas básicas.

¹¹ Un diseño operacional es un conjunto de etapas del proceso de planificación y conducción de la investigación, por medio de las hipótesis que se proponen como solución a un problema.

CAPÍTULO 3. EL MUESTREO

1. Introducción

Es común en la investigación social de tipo aplicada, la mayoría del conocimiento existente se debe a experiencias basadas en inferencias a partir de la observación y el análisis de un limitado número de unidades de análisis. De la calidad y representatividad de ese número limitado de unidades de análisis – muestra – dependerá la validez y confiabilidad de la información recolectada.

En los diseños cuantitativos, el proceso de escogencia de una muestra es importante. De allí la importancia de definir quienes van a ser estudiados (unidades de análisis), esto debe hacerse a partir de los objetivos del estudio.

Tanto el costo de cada entrevista personal como el límite de tiempo disponible no permiten realizar un censo o una cobertura total de la población de votantes, en estos casos, se recomienda usar muestras¹² cuando no es posible medir o contar a cada una de las unidades de análisis de la población¹³.

2. Objetivos

El objetivo final de una investigación social no se reduce solamente a la descripción de procesos de intervención, diagnósticos, evaluaciones sobre poblaciones particulares, en muchas ocasiones, lo que se pretende es generalizar los resultados de esos procesos. Matemáticamente es posible describir las características de una población a través de estadísticos¹⁴ muestrales.

¹² Los estadísticos usan el concepto *muestra* para describir una porción seleccionada de la población.

¹³ Generalmente se usa la palabra población no sólo para referirse a personas sino a todos los elementos que se han escogido para el estudio.

¹⁴ Un estadístico es una característica de una muestra representativa.

3. Desarrollo

El muestreo es la técnica con que se determina el tamaño y los elementos que integrarán la población muestral a fin de que cumpla la condición de ser representativa de toda la población. Hay varios tipos de muestreo y cada uno tiene diferentes fórmulas aritméticas para calcular el tamaño de la muestra, así como distinto método de elegir los integrantes de ésta.

3.1. Determinación de la muestra

Cuando se analizan todos los integrantes de una población se tiene un censo. Cuando se analizan algunos casos, pero que cumplen la condición de ser representativos de toda la población, se tiene una muestra poblacional. Ambos procedimientos tienen ventajas y desventajas. En el censo se tiene la ventaja de incluir a todos los miembros de la población, por lo que no hay necesidad de generalizar los resultados; la desventaja es que levantar un censo de población puede ser casi imposible por cuestiones de tiempo y recursos. La muestra tiene como ventaja que reduce la cantidad de recursos y tiempo requeridos; la desventaja principal está asociada con *el grado de representatividad de la información* recogida, la cual no olvidemos, se requiere para hacer inferencia al resto de la población.

A continuación, se muestra una tabla que contiene unos tamaños muestrales teóricos de acuerdo al grado de confianza requerido:

Tamaño de la muestra para un grado de confianza de:

	± 1%	± 2%	± 3%	± 4%	5%
Tamaño de la población					
500.000 y más	9423	2390	1065	600	3841.000
	$\frac{1}{2}$ ¹⁵	$\frac{1}{2}$	$\frac{1}{2}$	75	278
2.000	$\frac{1}{2}$	$\frac{1}{2}$	696	462	322
3.000	$\frac{1}{2}$	1344	787	500	341
4.000	$\frac{1}{2}$	1500	842	522	350
5.000	$\frac{1}{2}$	1622	879	536	357
10.000	4899	1936	964	566	370
20.000	6489	2144	1013	583	377
50.000	8057	2291	1045	593	381
100.000	8763	3245	1056	597	383

Fuente: Extraído de Apéndice estadístico de Sampling in Auditing, H. P. Hill; J. L. Roth y H. Arkin, New York: The Ronald Press, 1982.

3.2. Métodos de muestreo

Los métodos de selección de las unidades de análisis que componen una población muestral, pueden clasificarse en dos tipos: Muestreo aleatorio o probabilístico¹⁶ y muestreo no aleatorio.¹⁷

¹⁵ En estos casos más de la mitad de la población o universo se requeriría como población muestral.

¹⁶ Muestreo aleatorio es aquel método de selección de una población muestral que se basa en la teoría de la probabilidad; en él todos los elementos de la población tienen una probabilidad de ser escogidos en la muestra.

¹⁷ Muestreo no aleatorio es un método de selección de una muestra poblacional sin utilización de aleatoriedad. El criterio de escogencia está sustentado en el conocimiento de la población, los objetivos del estudio y el tipo de diseño metodológico requerido.

3.2.1. Muestreo aleatorio o probabilístico

3.2.1.1. Muestreo aleatorio probabilístico simple

Es aquel en que cada individuo dentro de la población tiene la misma probabilidad de ser seleccionado como integrante de la muestra. Así, en una población de tamaño N , cada individuo de la población total tiene una probabilidad de ser elegido de $1/N$. Los elementos de la muestra se eligen en forma aleatoria, por ejemplo:

Por ejemplo, se requiere seleccionar una muestra representativa de un grupo de 3000 estudiantes matriculados en la Facultad de Administración de Empresas. Hay dos maneras de hacerlo mecánicamente o de forma aleatoria:

- Por sorteo: se seleccionan escogiendo al azar (10%) un número predeterminado de estudiantes.
- Con el listado de estudiantes, se asigna un número a cada uno de ellos (001 a 3000) y se seleccionan 100 números de una tabla de números aleatorios.

3.2.1.2. Muestreo sistemático

En este método los elementos se escogen aplicando un criterio de selección preestablecido y uniforme, por ejemplo: en la selección de 300 estudiantes de un grupo de 3000 se puede utilizar el siguiente procedimiento:

Con el listado de 3000 estudiantes y una población muestral preestablecida de 300 personas, se seleccionan dentro de la población (3000) a un número de "n" elementos a



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

partir de un intervalo Q ($Q = N/n$) de selección. Por ejemplo: $Q = 3000/300 = 10$; este valor indica que de cada 10 personas será elegida una hasta completar el listado de 3000.

3.2.1.3. Muestreo estratificado

Para este tipo de muestreo, la población se divide en subpoblaciones, grupos o estratos que se presentan en la población con base en algún criterio o atributo, como: lugar de residencia; grupo de edad o etario; género, etc. Y se selecciona una muestra para cada subpoblación o estrato. Los elementos se escogen al azar dentro de cada subpoblación o estrato.

3.2.1.4. Muestreo estratificado proporcional

Este tipo de muestreo estratificado es un caso particular de muestreo estratificado en el cual se escoge en cada grupo el número de elementos que es proporcional al tamaño del estrato. Es el más utilizado, dado que las particularidades de la población –sexo, edad, nivel de escolaridad, nivel de ingresos, etc. - pueden influir en la determinación de los elementos que integran la muestra.

- Nota metodológica

En el ejemplo presentado, una muestra de 300 estudiantes, representa el 10% del grupo de 3000. Se quiere establecer una estratificación proporcional en función de su rendimiento académico, así, el estrato A para los estudiantes con un promedio entre 3.0 y 3.5; el estrato B para estudiantes con promedios que varíen entre 3.6 y 4.0; y el estrato C para los estudiantes con promedios mayores de 4.0.

Se sabe que el tamaño de la muestra -300- representa el 10% de la población. En este estudio, se quiere determinar el peso de cada estrato en que se dividió la población estudiantil, según su rendimiento académico. A la población de los 3 estratos se les aplica idéntico factor proporcional (10%).

Desagregada la población estudiantil del Programa de Administración de Empresas, de acuerdo a un criterio proporcional, la distribución es la siguiente:

	Población	Muestra
		10%
Estrato A	770	77
Estrato B	1500	150
Estrato C	730	73
	3000	300

- Ejemplificación

El siguiente diseño muestral se ha extraído del estudio Imágenes de identidad, integración y conflicto entre la comunidad estudiantil de la Universidad de Antioquia (Ruiz, Jaime, Aigner, Miguel, 1991:5).

Dada la suposición que se hizo en el estudio de la distribución de las características en la población y ante la decisión lógica de estudiar éstas en una muestra de estudiantes, se optó por el empleo de la metodología del muestreo estratificado. Para el efecto, los estudiantes fueron agrupados en tres estratos o Niveles de agrupamiento:

<i>Estrato 1 o nivel inicial:</i>	<i>Estudiantes en semestre 1^o, 2^o y 3^o</i>
<i>Estrato 2 o nivel intermedio:</i>	<i>Estudiantes en semestres 4^o, 5^o y 6^o</i>
<i>Estrato 3 o nivel terminal:</i>	<i>Estudiantes en semestres 7^o y más.</i>

El número de estudiantes en cada Nivel, así como la ponderación de cada uno, se presenta en la Tabla 3.1.

Tabla 3.1. Número de estudiantes por nivel y su ponderación

Nivel de Agrupamiento	número de estudiantes	ponderación nivel. wh
I	7179	0.49
II	4026	0.27
III	3559	0.24
Total	14756	1.00

Dado que para la investigación, los tres estratos presentan diferentes niveles de información, entonces se descartó el procedimiento más utilizado en este tipo de muestreo, a saber, la utilización de la división proporcional de la muestra total en los niveles de acuerdo con la ponderación de ellos. Si no, que por las consideraciones anteriores, se estableció el siguiente orden o clasificación: Nivel II, Nivel III y Nivel I. Los pesos muestrales asignados a éstos fueron: 0.5; 0.3 y 0.2. La siguiente tabla presenta la distribución de las ponderaciones por Nivel:

Tabla 3.2. Ponderación poblacional y muestral de los niveles

Nivel	ponderación wh	ponderación wh1
I	0.49	0.2
II	0.27	0.5
III	0.24	0.3

Luego del análisis de diversas alternativas, el grupo investigador decidió fijar como nivel de confiabilidad el 90% y un error máximo de 3.5% en la Estimación de la proporción, lo



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

cual a su vez representa un tamaño de muestra de 850 estudiantes, distribuidos por Nivel

de la siguiente manera:

170 para el Nivel I
450 para el Nivel II
255 para el Nivel III

Tabla 3.3. Tamaño de la muestra por nivel según ponderación

Nivel	Afijación Proporcional		Afijación Establecida		Afijación Real	
	Ponder.	Muestra	Ponder.	Muestra	Ponder.	Muestra
I	0.49	398	0.2	162	0.25	202
II	0.27	219	0.5	406	0.43	349
III	0.24	195	0.3	244	0.32	261

Como se anotó en la selección de estudiantes, la muestra total del nivel correspondiente (I-II-III) se repartió entre los diversos programas académicos en forma proporcional al peso de ellos en el nivel académico correspondiente."

3.2.1.5. Muestreo por conglomerados

En esta clase de muestreo, también se subdivide la población en grupos denominados conglomerados. La diferencia de este muestreo con respecto al *muestreo estratificado* es que se deben censar todos los integrantes de los conglomerados escogidos y no sólo algunos elegidos al azar. Los conglomerados se pueden seleccionar en forma aleatoria o con probabilidad proporcional al tamaño.

- Ejemplificación

Se cuenta con un listado de los 14.476 estudiantes matriculados en los programas académicos que ofreció la institución universitaria en el semestre I de 2003.

Supóngase que la lista está en orden alfabético y se seleccionan los estudiantes que integrarán la muestra por medio de una tabla de números aleatorios. Después se busca cada alumno en su facultad y programa académico correspondiente, se le asesora en el diligenciamiento del cuestionario, etc. Dado que este procedimiento es difícil y poco

práctico, resulta entonces más sencillo escoger algunos programas académicos y aplicar cuestionarios a todos los estudiantes. En este caso cada programa académico es *un conglomerado*.

- Nota metodológica

El tamaño de la muestra de la ejemplificación se obtuvo de esta manera:

a) El tamaño de muestra para proporciones está dado por:

$$n_o = \frac{P Q}{V}$$

Donde:

n_o = tamaño de la muestra

P = probabilidad de que se realice el evento

Q = probabilidad de que no se realice el evento

V = varianza del estimador

La varianza del estimador es:

$$V = \frac{d^2}{T^2}$$

Donde:

d = error permitido en los datos (está en función de la precisión deseada).

"t" = valor de la abscisa en el Eje X de una distribución normal tal que deje en la parte central un área igual a la confianza deseada.

$$n_o = \frac{t^2 P Q}{d^2}$$

b) Se establece que se requiere un 95% de confianza y que el error no sea mayor del 4%. Se busca en la Tabla del Área Bajo la Curva Normal y se identifica que "t" = 1.96, o sea, aproximadamente 2. Se estima que existe un 50% de probabilidad de que se efectúe el evento y, por lo tanto, un 50% de probabilidad de que no ocurra (Q = 100-P).

Se tiene al sustituir:

$$n_o = \frac{(4)(50)(50)}{16} = 625$$

c) Para poblaciones finitas se ajusta al tamaño de la muestra, pues en ocasiones resulta mayor que el de la población. El ajuste para poblaciones finitas está dado por:

$$n = \frac{n_o}{1 + \frac{n_o}{N}}$$

Donde:

n = tamaño ajustado de la muestra

n_o = tamaño no ajustado de la muestra

N = tamaño de la población

Entonces sustituyendo:

$$n = \frac{625}{1 + \frac{625}{14.756}} = 599,21$$



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

Tamaño de la muestra (n) = 599 estudiantes representativos de la población total estudiantil

Obsérvese que el cociente n_0 / N es crítico para tamaños pequeños de población; pero casi no tiene efecto en poblaciones grandes, como en este caso, donde la población es igual a 14.476 estudiantes.

3.2.2. Muestreo no aleatorio - no probabilístico

En éste tipo de muestreo, se utiliza el conocimiento y la opinión personal del investigador para identificar aquellos elementos de la población que deberán estar incluidos en la muestra.

3.2.2.1. Muestreo dirigido – no probabilístico

Tales muestras pretenden seleccionar unidades de análisis que de acuerdo al criterio del experto son representativos de una población en estudio.

3.2.2.2. Muestreo por cuotas

Se utiliza en estudios en donde el criterio de desagregación de la población son ciertas características de las variables demográficas de la población, por ejemplo, jóvenes profesionales menores de 30 años. El analista selecciona las personas con esas características hasta completar la cuota definida intencionalmente.

CAPÍTULO 4. LA RECOLECCIÓN DE LA INFORMACIÓN

1. Introducción

¿En que consiste la medición en la investigación social? Muchos diseños de investigación requieren la recolección sistemática de información sobre poblaciones totales o de sectores representativos (muestras poblacionales). Una vez seleccionado el diseño de investigación de acuerdo con el problema de estudio y sus hipótesis, el paso siguiente será recolectar la información pertinente.

2. Objetivos

Familiarizarse con la medición en la investigación social que utiliza preferentemente los diseños cuantitativos y los instrumentos de recolección de información primaria.

3. Desarrollo

El proceso de recolección de información implica las siguientes etapas:

- Diseñar o adaptar un instrumento válido y confiable de recolección de información.
- Definir una estrategia de recolección de la información
- Determinar la población informante –diseño de la muestra-.
- Realizar el trabajo de campo: aplicación del instrumento de recolección de información
- Validación y crítica de la información recolectada.
- Codificación y sistematización de los datos.
- Análisis e interpretación de la información

En este Capítulo se muestra, en términos generales, algunas consideraciones importantes para tener en cuenta en relación con los niveles y criterios de medición



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

utilizados en los diseños cuantitativos. Y también, ofrece una información general sobre los métodos más usuales de recolección de información.

3.1. La medición en los diseños cuantitativos

¿En que consiste la medición en la investigación social? En este campo ella se puede definir como el proceso por el cual una situación, hecho o comportamiento social es observado sistemáticamente y representado por puntajes alfanuméricos.

Comúnmente, se considera que las mediciones de las variables deben ser *objetivas* para que los resultados de la medición sean siempre los mismos, independientemente de quién los haga; y *cuantitativa*, en el sentido de usar puntajes numéricos para medir el comportamiento social.

3.1.1. Los niveles de medición y los estadísticos apropiados a cada nivel

Cuando se utiliza un diseño cuantitativo para desarrollar un proceso investigativo, nos encontramos con unos datos que requerirán, para su tratamiento y posterior análisis, diferentes niveles de medición. Es importante entender las características y diferencias de cada uno de ellos, ya que cada uno exige diferentes procedimientos estadísticos¹⁸. A diferencia de las ciencias naturales, en donde las convenciones de medidas son más precisas como, por ejemplo, el metro, el gramo, el litro, etc., en las ciencias sociales, las convenciones conceptuales denotan propiedades que no son generalmente tan precisas ni estructuradas, en su resultado.

- **El nivel nominal.** Este nivel de medición es simplemente un criterio de clasificación, no implica orden, ni jerarquía. Las unidades de análisis -individuos, regiones, propiedad, o situación o hechos- son simplemente clasificados en dos o más categorías. Por ejemplo, estado civil es una medida de nivel nominal que puede clasificar a las personas en

¹⁸ Generalmente, los niveles de medición nominal y ordinal utilizan técnicas no-paramétricas y los niveles intervalles o de razón utilizan técnicas paramétricas.

diversas categorías: casados, solteros, viudos, separados, unión libre, sin que tal clasificación implique un orden o una jerarquía.

El nivel ordinal. Este nivel indica, no solo un criterio clasificador, sino también un sentido de orden. En la medición ordinal se puede afirmar que la categoría X es más o menos (+ o -) que la categoría Y.

Supongamos que estamos preguntando a los estudiantes sus apreciaciones sobre el desempeño del Departamento de Bienestar Estudiantil y podría presentarse un continuo de respuestas que representa una escala que va de:

Muy de acuerdo - De acuerdo - No tiene criterio o no sabe - Desacuerdo - Muy en desacuerdo.

El conjunto de respuestas es un buen ejemplo de una medición ordinal ya que se clasifica y ordena a las personas que respondieron: Hay un orden entre los estudiantes que expresaron una actitud *Muy de acuerdo* con los programas asistenciales del Departamento de Bienestar Estudiantil con los que expresaron una actitud *Muy en desacuerdo*.

- Nota metodológica

Si se le asigna el número 1 a los que expresaron estar "Muy de acuerdo", el número 2 a los que están de "Acuerdo" y así hasta la última categoría, donde el número 5 se asigna a los que expresan el mayor desacuerdo, nos encontraremos tanto con un criterio de ordenamiento como también de clasificación.

Otro ejemplo de medidas ordinales son las escalas de actitudes¹⁹, en donde al encuestado se le presenta una lista o continuo de items o proposiciones relacionadas con la situación que se estudia, conjuntamente con un continuo de respuestas que representa las posibles valorizaciones que le asigna a la dimensión. La escala puede variar en un rango de: de mayor a menor, de acuerdo-desacuerdo, de satisfacción a insatisfacción.

¹⁹ "Una actitud es una organización relativamente duradera de creencias en torno a un objeto o situación las cuales predisponen a reaccionar preferentemente de una manera determinada". Rokeach. Milton. Naturaleza de las actitudes.

- Ejemplificación

Supongamos nuevamente que estamos interesados en medir las actitudes del medellinense sobre las causas de violencia, veamos que nos dice la siguiente tabla:

Tabla 1. Causas de aumento de la violencia

	Muy de Acuerdo	Acuerdo	En desacuerdo	Muy en desacuerdo	No Sabe N /R
El aumento de la violencia es un hecho normal	4. %	37.2%	46.5%	8.8%	2.9%
La violencia son acciones aisladas exageradas por la prensa	5.5%	48.6%	35.5%	4.6%	5.8%
La inseguridad y violencia son acciones de la delincuencia organizada	19.0%	57.6%	16.3%	2.4%	4.8% ³

- Nota metodológica

La tabla anterior nos muestra el peso porcentual de las opiniones²⁰ de un grupo de encuestados frente a unas afirmaciones que se escuchan a diario acerca de las posibles causas del aumento de la violencia en la ciudad de Medellín; ahora bien nos interesa individualizar estas actitudes ciudadanas y medir la intensidad. Supongamos que se le

²⁰ Distribución de frecuencia



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

presentan nuevamente los 3 ítems anteriores a un grupo de personas y se les pide su grado de acuerdo o desacuerdo con las afirmaciones anteriores.

A cada uno de los encuestados se les presentó tres proposiciones o declaraciones referidas al tema; cada una con un continuo de posibles respuestas escalares que representaban una escala ordinal:

1. Muy de acuerdo; 2. De acuerdo; 3. Indiferente; 4. En desacuerdo; 5. Muy en desacuerdo.

Al sumar el valor de cada una de las respuestas, una persona podría recibir desde 3 hasta 15 puntos: Las personas que manifestasen estar "muy en desacuerdo" con los planteamientos sobre las posibles causas del aumento de la violencia, podrían obtener un puntaje de máximo de 15 puntos; en cambio, aquellas que afirmasen, en todos los items, estar "muy en desacuerdo" obtendrían un puntaje de tres puntos.

De esta forma podemos *ordenar* en una escala a cada persona encuestada, desde los que estaban *muy de acuerdo* hasta *muy en desacuerdo*; así se logra construir *una escala ordinal*.

- Los niveles de medición de intervalo y de razón o proporción

Los niveles de medición de intervalo y de razón o proporción permiten hacer mediciones exactas de distancia entre dos puntos cuyos intervalos son iguales; las medidas de intervalo y razón generalmente son producto de convenciones operacionales tales como por ejemplo, peso y estatura de las personas; horas de capacitación u horas de trabajo semanal; nivel de ingresos; distribución de la población por grupos de edades; años de escolaridad formal, etc. Los niveles de medición de razón o proporción, además de reunir todas las características anteriores, vale decir, clasificar, ordenar, distribuir en rangos o intervalos iguales entre categorías, tienen como punto de partida el cero absoluto, el cual indica que en la escala hay un punto de partida sin la propiedad o característica que se está midiendo.

Las diferencias entre mediciones de intervalo y razón son de orden más bien práctico: las escalas de razón tienen un punto de partida que es el cero absoluto (0), situación que es común de encontrar en datos porcentuales o relativos o en datos proporcionales:

Las características de los niveles de medición se resumen en el siguiente cuadro sinóptico:

Niveles de Medición	Características
Nominal	Sólo clasifica.
Ordinal	Clasifica y ordena según magnitud. (Más que o menos que
Intervalo	Clasifica, ordena en rangos, estratos o intervalos.
Razón	Su punto de partida es cero (0), lo cual indica la
.	Clasifica, ordena según magnitudes en rangos iguales y ausencia de la propiedad medible

A continuación se hará una presentación de los estadísticos y técnicas de análisis más apropiadas para cada uno de estos niveles de medición.

- Los niveles de medición y las estadísticas apropiadas a cada nivel (Siegel, 1978:51)

Escala	Relaciones definidas	pruebas estadísticas apropiadas	Tipos de estadísticas apropiadas.
Nominal	Equivalencia	Pruebas estadísticas no paramétricas	Moda Distribución de frecuencia Coef. de Contingencia
Ordinal	Equivalencia de mayor a menor		Mediana Percentiles Rho de Spearman
Intervalo	Equivalencia de mayor a menor	Prueba estadísticas paramétricas	Media aritmética Desviación estándar
Proporción	Proporción Continua		Correlación del Pearson Regresión múltiple Análisis multivariado, etc.

3.1.2 Conceptos básicos de medición

La medición se puede hacer en cuatro niveles, de acuerdo con la forma como los números – cantidades que se asignan a los objetos o individuos – se relacionan con la propiedad o característica que está siendo medida. Estos están íntimamente asociados con la presencia acumulada de las siguientes cuatro características: clasificación, ordenamiento, intervalos iguales y cero absoluto o continuidad.

3.1.2.1. Características de los niveles de medición

Características	Nominal	Ordinal	Intervalo	Razón
Clasificación	Si	Si	Si	Si
Ordenamiento	No	Si	Si	Si
Intervalos iguales	No	No	Si	Si
Cero absoluto	No	No	No	Si

- Ejemplificación

Si Ud. tuviera que clasificar la afiliación partidista de los medellinenses utilizaría una escala:

Nominal Ordinal _____
Intervalo _____ Razón _____

Un nivel de medición que permita clasificar y ordenar las características socioeconómicas (lugar de residencia, nivel de ingresos familiares, estratificación socio-económica, nivel de escolaridad de una población, como por ejemplo, los estudiantes de la Universidad Complutense de Madrid corresponde a una escala:

Nominal _____ Ordinal
Intervalo _____ Razón _____

Si se quiere medir la edad en años cumplidos, de los estudiantes de la Facultad de Educación y Ciencias Sociales, ¿qué escala será la más adecuada? :

Nominal _____ Ordinal _____
Intervalo _____ Razón

Si se quiere medir la población de mujeres en edad escolar agrupada en grupos etarios o de edades iguales (quinquenales), ¿qué escala será la más adecuada?:

Nominal _____ Ordinal _____
Intervalo Razón _____

Actitud hacia las políticas de bienestar estudiantil en Universidad.



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

1. Muy de acuerdo 2. De acuerdo 3. No sabe 4. En desacuerdo 5. Muy en
desacuerdo

Nominal _____
Intervalo _____

Ordinal **X**
Razón _____

3.1.2.2 Criterios de medición

La calidad interna y externa de la información depende en gran parte de los criterios de validez, confiabilidad, representatividad, comprensibilidad y totalidad conceptual (operacionalidad conceptual); Criterios que deben ser tenidos en cuenta por el analista social, desde el momento en que identifica un problema hasta el proceso de análisis e interpretación de la información obtenida.

- **La validez.** Es el grado en que un instrumento de medición realmente mide lo que debe medir. Un importante criterio para evaluar la calidad de un cuestionario radica en que lo que éste mida sea válido. Por ejemplo, si se quiere medir la opinión sobre el pacto social de la Consejería Presidencial para Medellín, hay que estar seguro que lo que se está midiendo es realmente la opinión ciudadana acerca de los planes y programas de dicho Programa y no una opinión abstracta o global formada o, por la prensa o, por sectores políticos o intelectuales.

¿Pero cómo uno puede saber si la medición es realmente válida? En la práctica no hay en las ciencias sociales un método único y confiable que permita concluir fehacientemente que tal dimensión conceptual ha sido realmente medida; Hay unos criterios generales que más se acercan a lo intuitivo que a lo científico, por ejemplo, la técnica de validez de contenido que exige que se abarquen todos los elementos o características posibles que se suponen es parte de la dimensión conceptual que se pretende medir.

Este criterio metodológico de validez se puede apreciar cuando en un sondeo se pretende medir la opinión de los docentes de la Universidad de Antioquia respecto al escalafón docente. Un sondeo de opinión como el planteado no se puede reducir a medir

sólo las apreciaciones acerca del recorrido y ascenso a través de los años o sólo como un problema de méritos, se tendría necesariamente que abarcar otras dimensiones tales como: producción intelectual e investigativa, dedicación y experiencia profesional; capacitaciones y especializaciones logrados, etc.

- **Confiabilidad.** Una medida confiable es aquella, que aplicada sucesivamente a través del tiempo, producirá idénticos resultados. (Se asume que el objeto medible no cambia a través del tiempo). El criterio de confiabilidad es fuente continua de discusión y de sesgo en la investigación social; básicamente se manifiesta por la utilización de preguntas vagas, complejas o no pertinentes con la temática que se está estudiando o porque los encuestados no logran entender la pregunta; Aún las preguntas más sencillas pueden conllevar cierto grado de indeterminación y de desconfianza en las respuestas obtenidas.

*¿Cómo decide si una medida es confiable? Varias pruebas han sido desarrolladas para este fin. Una de las más utilizadas es el método de **prueba re-prueba**. Una medición confiable dará los mismos resultados cuando se aplica sucesivamente en diferentes tiempos, suponiendo que no se den cambios en el objeto-sujeto que se está midiendo.*

¿Cómo podemos distinguir los criterios de validez y confiabilidad? Unos ejemplos nos pueden ayudar a clarificar uno de otro.

- Ejemplificación

En un estudio sobre la obra de equipamiento urbano del Instituto Metropolitano de Valorización que comprende la ampliación de la frontera urbana (ampliación, apertura y amoblamiento de vías) en la zona suroriental por el sistema de

valorización; se utilizó un diseño tipo panel, para evaluar las características socio-económicas de los residentes.

Entre quienes fueron inquiridos con el mismo cuestionario cuatro meses más tarde, un porcentaje apreciable mostraron inconsistencias en sus respuestas sobre la edad declarada; el ingreso percibido; número de personas a cargo y bienes declarados. Si bien la mayoría se mantuvo consistente en sus respuestas, esto muestra que la confiabilidad de la información obtenida en los llamados estudios de opinión es algo que hay que controlar o tener presente (Ruiz, (1999), p....)

- **La comprensibilidad.** Un analista generalmente está escribiendo informes para un lector que no es especialista en el tema, por lo tanto, es muy importante escoger conceptos e indicadores que sean entendibles por el ciudadano común. Medidas sobre conceptos o indicadores demasiados complejos, no importan que tan confiables y válidos sean, no serán entendidos o tomados en cuenta.

- **Totalidad conceptual.** Es otra característica de una buena medición. Volvamos al ejemplo del estudio sobre seguridad comunitaria discutido anteriormente; obviamente la escala de 5 ítems o enunciados, nos dio mejor información acerca de la actitud de los residentes con respecto al programa, que un cuestionario de una o dos preguntas. De igual forma la utilización de múltiples enunciados facilita la posibilidad metodológica de abarcar el máximo posible de apreciaciones, elementos o características de una dimensión.

- **Ejemplificación**

Se está interesado en estudiar la apreciación de los estudiantes sobre los espacios verdes de la ciudad universitaria. Una pregunta, como por ejemplo, *¿Qué opina sobre el estado de los espacios verdes en la Ciudad Universitaria?*, A un nivel de generalidad tal, solo determinaría confusión porque algunos encuestados podrían pensar que se les pregunta sobre el uso de esos espacios, otros acerca del estado físico de esos espacios o de su mantenimiento. Lo anterior es un buen ejemplo de la necesidad de diseñar las preguntas de tal forma que realmente *inquieran acerca de lo que realmente se quiere preguntar*.

- Nota metodológica

En el caso específico de la ejemplificación anterior muestra que hay que preguntar acerca de dimensiones específicas del entorno de estos espacios verdes en los cuales el estudio esta interesado; esto traerá ventajas también en los resultados, ya que aportará información específica. Lo recomendable sería diseñar preguntar acerca de dimensiones tales como: el estado o mantenimiento de los espacios verdes; la seguridad en ellos; el ornato; los accesos, la infraestructura disponible - asientos, mesas, etc. En síntesis, las respuestas que uno obtenga darán cuenta en forma más integral o completa de las dimensiones que se quieren analizar, si estas abarcan la totalidad de dimensiones del problema de investigación.

3.2. Instrumentos de recolección de información

En el proceso de investigación, la etapa de recolección de información ocupa un lugar básico. Pero previo a esta etapa debemos resolver que tipo de información es la más conveniente para comprobar nuestras hipótesis de trabajo.

Un estudiado planteamiento de un problema de investigación y un enunciado claro y conciso de la(s) hipótesis nos va a definir entre otras decisiones metodológicas:

- ¿Qué tipo de información requerimos?
- ¿Cuándo necesitamos construir nuestros propios instrumentos de recolección de información y cuando no?
- ¿Cuándo nuestra hipótesis de trabajo exige un alto nivel probatorio y cuando uno bajo?

En algunas ocasiones se debe utilizar información primaria es decir, aquella que el investigador recoge directamente a través de un contacto inmediato con su objeto de análisis. Y en otras ocasiones información secundaria que es "aquella que el investigador recoge a partir de investigaciones ya hechas por otros investigadores con propósitos diferentes.

La distinción entre estos dos tipos de información, nos señalará la técnica para recolectar el tipo de datos requeridos:

Técnicas de recolección de Información primaria

La encuesta social

técnicas de recolección información secundaria

Análisis de contenido documental



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

La entrevista social

La observación científica

La experimentación

Análisis de medios audiovisuales

Análisis de datos agregados

3.2.1 Conceptos básicos de la recolección de información

El proceso de selección de *la información necesaria y pertinente* no es estrictamente un proceso intuitivo, al menos no lo es en los diseños cuantitativos. Las dimensiones conceptuales utilizadas en la formulación del problema y las hipótesis están señalando qué tipo de información se requiere. En términos metodológicos, se puede afirmar, que existen unos lineamientos básicos para recolectar la información –datos- y se sintetizan en lo que se denomina una estructura básica común a cualquier tipo de datos y es la siguiente:

La estructura tripartita:

Unidad de análisis	Variable	Valores de la variable
--------------------	----------	------------------------

- Nota metodológica

En el estudio ya citado sobre *las apreciaciones de los estudiantes sobre su mundo universitario*, es posible apreciar desde el planteamiento del problema, cuál es el sujeto-objeto del estudio; las hipótesis que explícita o implícitamente se plantean; qué tipo de información será pertinente; cuáles son las dimensiones principales del estudio. Analicemos apartes de la Introducción del citado estudio que nos indica estos tres elementos que componen la estructura tripartita del dato:

"Una gran parte de los estudios referidos a los estudiantes universitarios ha puesto su atención, bien, en la descripción de las características sociales, económicas y culturales de los estudiantes y su entorno familiar, o bien, en las manifestaciones del movimiento estudiantil. Estas perspectivas de análisis, comúnmente han renunciado al análisis de esos aspectos que para la investigación sociológica son importantes, a saber las

prácticas estudiantiles (académicas, existenciales, institucionales universitarias, institucionales sociales) frente a la vida universitaria, sus potencialidades de integración y de conflicto. (Ruiz, Jaime, et. al., 1991:45).

- *La unidad de análisis* en este estudio es el estudiante de pregrado matriculado en programas académicos presencial que ofrece la Universidad de Antioquia en el semestre 2 de 2003. En la ejemplificación anterior es posible apreciar lo que aquí se denomina *la estructura tripartita del dato*. Primero, se identifica un sujeto-objeto del estudio, en este caso son los estudiantes de la institución. Pero, *¿Qué es lo que se quiere saber de esta unidad de análisis?*

El principio básico del diseño de estudio de opinión que presentamos, es la descripción de estos elementos, relaciones, situaciones, prácticas sociales, académicas y existenciales que se manifiestan en el contexto de la Universidad de Antioquia.

Se ha definido que la unidad de análisis es el estudiante matriculado en programas de pregrado en la Universidad de Antioquia. El segundo paso es definir una serie de dimensiones o variables o características asociadas a la unidad de análisis (alumno universitario) que interesa conocer:

- Aspectos sociodemográficos y académicos generales de la población estudiantil
- Las imágenes que proyecta la universidad
- Factores de integración y conflicto
- La vida cotidiana estudiantil en la Universidad.
- Los niveles de participación del estudiante



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

Finalmente, debemos definir *los valores de las variables*, es decir, qué queremos saber de esas dimensiones detalladas anteriormente.

Resumidamente, el formato de la estructura de los datos en el estudio sobre los estudiantes de la Universidad de Antioquia puede ser el siguiente, en algunas de las variables consideradas:

Unidad de análisis	VARIABLES	Valores de las variables
Estudiante	Composición de la población por sexo	Masculino Femenino
	Estado civil de los estudiantes soltero casado separado viudo Unión libre	
	Horas de estudio diario	< 1 1 - 2 3 - 4 4 o más

- Ejemplificación

Veamos en la siguiente tabla, el formato que adquiere la estructura tripartita del dato, en una situación real:

- Se sabe que la unidad de análisis son los estudiantes
- En éste caso específico, *la variable* hace referencia a la percepción que se tiene sobre los problemas que vive la Universidad.
- Los valores que asume la variable serán las diferentes categorías de respuestas percibidas con su respectivo peso porcentual dentro de las opiniones.

Tabla 4.2. El principal problema en el interior de la ciudad universitaria y fuera de ella²¹

Problema	Ciudad Universitaria	Fuera de Ciudad Universitaria
Inseguridad interna	24.2%	33.0%
Deficiente dotación de bibliotecas y centros de documentación	4.2	2.4
Deficientes ayudas audiovisuales	2.0	1.2
Bajo nivel del profesorado	15.3	10.0
Exceso de profesores de cátedra	9.6	4.1
Débil vocación del estudiante	9.0	12.3
Deficientes servicios asistenciales	4.2	6.2
Poca oferta de cupos	10.0	9.3
Deficiente comunicación directivas- Estudiantes	17.0	16.4
Otros	4.2	5.0

²¹ Ruiz. R., Jaime, Aigner. Miguel., et. al. *Ibíd.*, p. 94.

3.2.2. Técnicas de recolección de información

A continuación presentamos algunas de las técnicas de recolección de información más utilizadas en los diseños cuantitativos porque facilitan el manejo de la información cuantificable.

3.2.2.1. Medición escalar

Existe un tipo de información que por sus características –subjetivas, del orden de la opinión, de las imágenes, de los valores- no permite ser medida en su cabalidad con instrumentos tradicionales; en estos casos, *la medición escalar* coloca al encuestado en posibilidad de expresar su opinión dentro de una gradación que mide la intensidad de la opinión. Esta técnica permite una mayor aprehensión de lo cualitativo del hecho social que se estudia.

Una escala es una técnica que permite medir *la cantidad de una propiedad, llamada actitud hacia algo, poseída por un conjunto de personas, (Escalante, Carlos, 1983:11)*

Se dice que una medición es integral, cuando ella abarca adecuadamente todos los aspectos del concepto en estudio. A menudo, esto es difícil de realizar con una sola medida, tal como con una pregunta en un cuestionario.

- Nota metodológica

En muchas circunstancias, un investigador querrá obtener varias mediciones y combinarlas en un índice o en una escala. Veamos el diseño de una escala que permita evaluar el desempeño académico de los estudiantes del Programa de Sociología de la



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

Universidad de Antioquia, medido por su mayor o menor dedicación a un conjunto de actividades que suponemos determinan su rendimiento académico.

ceo@carios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

¿Cuánto tiempo le dedica usted a las siguientes actividades relacionadas con su desempeño académico?

Objeto de actitud medido	Afirmaciones		
	Siempre -	Ocasional -	Nunca -
Desempeño académico -	3	2	1
Valores			
Estudiar notas de clases	{ }	{ }	{ }
Revisar la bibliografía	{ }	{ }	{ }
Asistencia a clases	{ }	{ }	{ }
Consulta textos	{ }	{ }	{ }
Trabajo en equipo	{ }	{ }	{ }
Consulta al profesor	{ }	{ }	{ }
Totales			

Una matriz de respuestas, en la cual cada alternativa de respuesta tiene un valor o peso dentro de la escala, permite ordenar a los encuestados en un índice o jerarquía. En el ejemplo anterior se puede hacer una "escala" que va desde 6 puntos –desde el estudiante que declara que *nunca* realiza las actividades enumeradas- hasta aquél que reúne 18 puntos, lo cual sería el caso del estudiante que afirma realizar *siempre* todas las actividades.

La escala²² más aconsejable es aquella que presenta *afirmaciones* que deben ser evaluadas en un continuo de categorías de no más de 5 valores.

- La selección de las afirmaciones o proposiciones de una escala

²² "Se puede definir una escala como un conjunto de ítems o declaraciones contruidos por el investigador según ciertas definiciones conceptuales previas y utilizadas para cuantificar características o propiedades del comportamiento." (Escalante, Carlos, 1983:12)

En esta primera etapa del diseño de la escala hay que poner especial cuidado en la validez y relación de *las afirmaciones*, (Hernández, R, et, al., 1991:264) que se van a utilizar para medir la dimensión conceptual del estudio, ya que si un aspecto importante es omitido, la escala puede ser considerada inválida²³.

La mejor estrategia para probar si existen *omisiones* en las afirmaciones o proposiciones sobre la dimensión del estudio es incluir la categoría *otras y especifique cuáles* en la encuesta piloto; aquellos valores más frecuentemente mencionados deberán ser incluidos en el cuestionario final.

- Interrelación entre las afirmaciones

Si todas las afirmaciones tienen como objetivo medir el mismo concepto, ellas deben estar relacionadas una a una y en conjunto. La escala debe ser unidimensional, esto es, medir una sola dimensión del hecho, proceso, actitud o conducta que se está investigando.

En el caso del modelo de escala utilizada para medir la intensidad de tiempo que el estudiante le dedica a un conjunto de actividades que se relacionan con su desempeño académico, se está tratando de diferenciar entre estudiantes *que siempre realizan una actividad* con aquellos que *nunca las realizan*; quienes logran un puntaje alto (3 puntos) en dos o más respuestas deberían obtener semejantes puntajes en las respuestas siguientes con algunas excepciones, si las tendencias muestran una dispersión en las respuestas se están midiendo dos o más dimensiones del desempeño académico.

- Valorización de la escala.

²³ "Las afirmaciones califican al objeto de la actitud que se está midiendo", (Hernández, R, et. al. 1991:264)

Un tercer paso es asignar un puntaje numérico a cada respuesta obtenida. Esto implica decidir acerca del valor de cada respuesta y luego sumarlas. A continuación, las respuestas deberán ser agrupadas en algunas pocas categorías, tales como *alta*, *media* o *baja*. Al examinar el total de los puntajes, usualmente se descubre que los extremos – altos o bajos, siempre o nunca – son escasos.

- Análisis de los de los resultados obtenidos mediante una escala

La lectura de los datos obtenidos puede ser un ejercicio aritmético de sumatoria de los puntajes obtenidos por cada uno de los respondientes.

- Algunas clases de escalas:

- **Escala de Likert.** Esta es una técnica de medición multidimensional²⁴ que trabaja generalmente con datos a nivel ordinal²⁵ y que permite medir una actitud y el comportamiento consecuente con ésta. Se basa en la premisa según la cual *la intensidad de las apreciaciones, opiniones o imágenes de un individuo* puede ser escalada dentro de ciertas categorías. Así, en lugar de preguntar abiertamente a las personas su opinión acerca de un tema, ellas son enfrentadas a unos enunciados o afirmaciones cuyos contenidos, por sus características, los colocan ante la posibilidad de tener que manifestar sus opiniones dentro de una escala de categorías posibles de respuestas; y la variación de tales categorías puede abarcar, por ejemplo, un rango desde *muy de acuerdo*, *acuerdo*, *indiferente*, *desacuerdo*, hasta *muy en desacuerdo*. Un aspecto distintivo de la escala de Likert es que asume que *las afirmaciones* miden la actitud hacia *un único objeto – concepto subyacente*. El procedimiento para diseñar y codificar una

²⁴ Una característica propia de la escala Likert es que no es unidimensional; los enunciados o afirmaciones o ítems no necesariamente deben estar relacionados temáticamente unos a otros.

²⁵ Al igual que *la escala de Guttman*, se trabaja con datos ordinales; existen escalas de medición como *la escala de Thurstone* que pueden producir mediciones a nivel de intervalos.

escala tipo Likert, es similar a las que hemos presentado anteriormente: debe construirse, al menos, *con 10 afirmaciones que califiquen el objeto de actitud.*

Ejemplo de una escala Likert aplicada a una pregunta para seleccionar proveedores, *¿qué tan importante para su empresa es...?*

	Muy Importante (5)	Importante (4)	medianamente importante (3)	- Poco importante (2)	- No es importante (1)
Calidad producto	X				
Precio	X				
Forma pago		X			
Tiempo entrega			X		
Garantía		X			
Marca del producto			X		
Prestigio marca		X			

Las respuestas se califican de acuerdo a los valores preestablecidos y las puntuaciones de las escalas Likert se obtiene sumando los valores obtenidos respecto a cada *frase o afirmación*. Por ejemplo, el valor obtenido es: $5 + 5 + 4 + 3 + 4 + 4 + 4 = 29$

Una puntuación se considera alta o baja según el número de *afirmaciones*. Por ejemplo, en la escala para evaluar la actitud para seleccionar un proveedor para la empresa la puntuación mínima posible es de 7 y máxima es de 35. La persona de la ejemplificación obtuvo 29 puntos, se puede concluir que su actitud hacia los proveedores es más bien favorable.

- Nota metodológica

En el estudio de opinión *Imágenes de identidad, integración y conflicto entre la comunidad estudiantil de la Universidad de Antioquia*, (Ruiz, Jaime, et. al., 1991: 21), se

les pidió a los estudiantes su valoración, acerca de lo que el estudio denomina *factores integradores*, en términos de las categorías *adecuadas, inadecuadas e indiferentes*;

La tabla siguiente muestra los resultados obtenidos, en el mencionado estudio mediante un formato escalar:

Tabla 4.3. Valoración de las prácticas existenciales dentro y fuera de la Ciudad Universitaria²⁶

Actividades	Adecuadas		Inadecuadas		Indiferentes	
	C.U.	F.U.	C.U.	F.U.	C.U.	F.U.
Teatro	91.0%	89.1%	2.3%	5.4%	6.8%	1.0%
Conciertos	88.3	88.8	3.6	2.0	8.1	1.0
Deportivas	53.1	57.65	14.7	11.7	2.2	30.7
Recreativas	54.2	54.1	13.7	15.6	31.1	30.3
Cine	9.1	16.6	61.9	46.3	29.0	37.0

C.U. = En el interior de la Ciudad Universitaria. **-F.U.** = Por fuera de la Ciudad Universitaria.

²⁶ Ibíd. p. 24

- Escala de Guttman

Esta es una técnica que presenta múltiples escogencias a lo largo de una dimensión. El propósito de esta escala es medir la *unidimensionalidad actitudinal* -mide sólo una dimensión- de acuerdo con el supuesto de que la actitud integra está contenida en una sola dimensión. Es decir, los enunciados o afirmaciones deben estar relacionados unos con otros. Los encuestados son enfrentados a responder a un conjunto de opciones o alternativas que se pueden dar en la dimensión actitudinal. Estas opciones se presentan en una especial disposición, de tal manera que las alternativas o preguntas, midan la intensidad de la apreciación o la opinión.

Supongamos que estamos midiendo actitudes frente a las prácticas abortivas. Para estos efectos se diseñó un continuo de items relacionados con el tema:

¿Está usted de acuerdo o en desacuerdo con que el aborto sea permitido en las siguientes circunstancias? :

1. Cuando la vida de la mujer esta en peligro
2. En caso de incesto o violación
3. Cuando el feto muestra clínicamente riesgos graves de salud
4. Cuando hay consentimiento paterno
5. Cuando la mujer no puede criar responsablemente el futuro niño
6. Cuando la madre lo considere pertinente.

Veamos algunos resultados obtenidos con esta pregunta en formato escalar:

Tabla 4.4. Actitudes frente al aborto

Persona	Riesgo para mujer	Incesto violación	Mal forma- Fetal	Consentí- miento paterno	Irresponsa- bilidad materna	Decisión mujer	Nº respuestas de acuerdo
1	A	A	A	A	A	A	6
2	A	A	A	D	A	D	4
3	A	A	A	D	D	D	3

ceo@catios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

4	A	A	D	A	D	D	3
5	A	D	D	A	D	D	2
6	D	A	D	D	D	D	1

Convenciones: *A = Acuerdo* *D = Desacuerdo*

3.2.2.2. La entrevista social

En algunas ocasiones, la información que se requiere, ya sea por sus características - complejidad -creencias, valores- o, por razones de orden psicológico o ideológico e incluso histórico, sólo es posible reconstruirla mediante un proceso conversacional. Si bien, este tipo de técnica no permite, en términos generales, recolectar una información factible de cuantificar, se convierte en una útil fuente de referencia, sustentación y complementación en los diseños cuantitativos.

El instrumento básico de esta técnica de recolección primaria es *la guía de entrevista o guión* diseñado teniendo como base las hipótesis o temáticas preestablecidas, lo cual permite obtener una información válida dentro de sus limitantes con la ayuda de replanteamientos, aclaraciones o ayudas audiovisuales, y a la vez, esto facilitará un registro más confiable de las respuestas.

3.2.2.3. La encuesta social

Esta es una técnica destinada a obtener información primaria de un grupo representativo de individuos para proyectar los resultados sobre la totalidad de la población en estudio, con el objeto de medir de la forma más acertada los conocimientos, las actitudes, las creencias y los comportamientos de las personas. La posibilidad de lograr lo anterior depende básicamente de la precisión, pertinencia y concisión de las preguntas.

El cuestionario es el instrumento básico de recolección de información de la Encuesta Social; consiste en un formulario con un conjunto de preguntas estandarizadas aplicadas a determinados sujetos o unidades de análisis.

- Los elementos constitutivos de la encuesta social

- Planteamiento de un problema de investigación
- Determinación de las hipótesis de trabajo y/o objetivos
- Definición y operacionalización de los conceptos y variables
- Diseño del cuestionario
- Diseño muestral
- Trabajo de campo
 - Pre-test o prueba piloto
 - Recolección, crítica y validación de la información.
- Sistematización
- Análisis e interpretación de la información
- Informe final.

- Temáticas propias de la encuesta social

- Características de la muestra poblacional: composición social, composición del hogar, acceso a medios de consumo colectivo, características demográficas;
- Opinión pública: actitudes, comportamientos e ideologías
- Conocimientos; tendencias electorales, mercadeo social y económico, etc.

- Características metodológicas de la encuesta social

1. Permite mayor cobertura en la población del estudio y mayor representatividad de ésta. Los resultados se pueden generalizar, gracias a las técnicas de muestreo y la inferencia estadística.
2. Hace posible traducir la información al lenguaje de variables-preguntas; y así se puede codificar la información mediante un sistema alfanumérico, lo que facilita la cuantificación.



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

3. Es una técnica complementaria a la Observación Social y la Entrevista Social.

ceo@carios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

- Diferencias entre las técnicas de la encuesta social, la observación y la entrevista

	Observación /Entrevista	Encuesta
Nº de unidades de análisis	Limitado	Ilimitado
Nº dimensiones variables del estudio	Limitado	Ilimitado
Representatividad	Menor	Mayor
Precisión Cuantificable	Menor	Mayor

- Proceso de diseño de un cuestionario

La base de las preguntas de un cuestionario son los conceptos y variables mediante los cuales se ha definido un problema de investigación. Pero, para pasar de los conceptos y variables a las preguntas se necesita una guía -un indicador- de lo que se debe observar.

El esquema es el siguiente:

Concepto	Variable	Dimensiones	Indicadores	Pregunta
----------	----------	-------------	-------------	----------

El proceso de convertir una dimensión (aspectos relevantes de una variable) en indicadores²⁷ se le denomina *operacionalización*²⁸.

Ejemplificación

Problema de investigación: Los estudiantes de la maestría en Administración de Empresas de la Universidad del Valle y sus prácticas existenciales. Gran parte de los estudios referidos a los estudiantes universitarios han puesto su atención, en la descripción de las características sociales, económicas y demográficas de los estudiantes y de sus grupos familiares, o en las manifestaciones del movimiento estudiantil. Estas perspectivas de análisis, comúnmente han renunciado al análisis de un aspecto que para la investigación social es importante, como por ejemplo, las llamadas prácticas estudiantiles -académicas, existenciales, institucionales- frente a la vida universitaria.

²⁷ El indicador hace medible y observables, conceptos o variables, no directamente medibles y observables.

²⁸ Traducción de conceptos o ideas a términos cuyos referentes son observables, empíricos y medibles.

Objetivos de la investigación: Identificar las manifestaciones más corrientes de estas prácticas existenciales. Este análisis involucra una doble visión: las prácticas mismas y, la caracterización social de los actores.

Diseño del cuestionario: A título de ejemplificación vamos a trabajar el concepto *prácticas existenciales*. Del mismo planteamiento del problema, extraemos su definición (operacional), son *aquellas que tienen que ver con la existencia individual del estudiante, con su vida cotidiana, con las dimensiones no-universitarias de su existencia*. Esta definición nos sugiere varias dimensiones a analizar, tales como el proceso de reproducción social del estudiante; características socio-demográficas; vida cultural; utilización del tiempo libre -recreación- y composición social, etc. Para simplificar vamos solo a *describir* las características socio-demográficas del estudiante de la Maestría en Administración de Empresas de la Universidad del Valle.

Conceptos operacionalizados	Indicadores ²⁹	preguntas
- Distribución por sexo	Masculino Femenino	consignar sexo
- Distribución por edad	15 – 19 20 - 24 25 - 29 30 y más	¿Qué edad tiene Ud?
- Estado civil	soltero casado separado viudo Unión libre	Cual es su estado civil?

²⁹ La primera tarea es idear y construir "indicadores observables" a partir de los conceptos planteados anteriormente en el planteamiento del problema.

En el ejemplo anterior, se puede observar el proceso que va desde **la definición y operacionalización de los conceptos**³⁰, el diseño de los indicadores y las preguntas – es decir, la construcción del cuestionario-.

Diseño de la muestra -poblacional del estudio: selección aleatoria o probabilística de una población muestral

Trabajo de campo: Recolección de información mediante la aplicación de un cuestionario estandarizado

Sistematización de la información:

- Crítica y validación de la información
- Codificación: Libro de códigos y hojas de codificación
- Sistematización de la información

Análisis e interpretación de la información

Lectura y explicación de los resultados encontrados

Informe final:

Contenido técnico científico del proyecto

Contenido administrativo del proyecto

³⁰ Definiciones operacionales: son proposiciones que especifican un procedimiento para determinar el valor numérico de una variable dada en casos concretos. Es traducir un concepto-variable en indicadores.

- El cuestionario y las preguntas

El instrumento básico de recolección de información en la técnica de la encuesta social es el cuestionario, el cual consiste básicamente en un formato con preguntas y en ocasiones incluye las posibles respuestas. Con las preguntas del cuestionario se busca obtener información acerca de lo que una persona conoce, cree o espera, siente o desea, intenta hacer, hace o ha hecho.

Correlativas a las preguntas están las respuestas y sobre éstas, tradicionalmente, la discusión se centra en su *calidad externa e interna*. La calidad externa se refiere a la posibilidad de generalizar los resultados encontrados. En cambio, *la calidad interna* de la información recolectada se refiere a dos situaciones: *la validez y la confiabilidad*. La primera, procura que las preguntas reflejen realmente lo que se quiere medir; y la segunda, indica que un cuestionario o una pregunta es confiable cuando se aplica repetidamente con la misma metodología, en idéntica población produciendo iguales resultados.

- Tipos de preguntas

Los cuestionarios pueden presentar dos tipos básicos de preguntas: según el contenido y según la forma. Según *su contenido*, se pueden distinguir 2 tipos:

Preguntas relacionadas con hechos, susceptibles de confirmarse, mediante la comparación o valiéndose de información secundaria considerada válida o también, partiendo de situaciones vividas por el interlocutor, por ejemplo:

¿Ha consumido usted, por lo menos una vez en su vida, por lo menos una vez en su vida, alguna droga adictiva como marihuana o cocaína?

Preguntas de opinión: éstas tienen que ver con las opiniones, actitudes, preferencias del encuestado, por ejemplo:

¿Durante 2003 habrá en Medellín menos o más violencia que durante 2002?

Según *su forma*, podemos hablar de *preguntas abiertas o cerradas*, o de una mezcla de ambas. Las preguntas cerradas dan al encuestado la oportunidad de escoger una respuesta entre una serie de opciones o categorías preestablecidas; mientras que las preguntas abiertas no proponen respuestas posibles y dejan espacio para que cada cual responda, por ejemplo:

¿Con quién vives? _____

- Preguntas abiertas y cerradas: algunos limitantes

Un ejemplo de cuestionario con preguntas cerradas, con alternativas de respuesta preestablecidas y codificadas en el mismo instrumento sería:

¿Con quién vives?

- | | |
|------------------------------|--------|
| 1. Con tus padres y hermanos | 1. --- |
| 2. Con otros parientes | 2. --- |
| 3. Con hermanos mayores | 3. --- |
| 4. En una residencia | 4. --- |
| 5. Con tu esposa(o) | 5. --- |
| 6. Con amigos | 6. --- |
| 7. Solo | 7. --- |

En términos generales, la ventaja de utilizar preguntas cerradas en un cuestionario radica en la facilidad de codificación y comparabilidad. Otra ventaja de este tipo de preguntas en temas como participación política, frecuencia de la actividad sexual, ingresos, edad, es que los encuestados prefieren responder escogiendo una de las alternativas ofrecidas antes que tener que expresar verbalmente respuestas incómodas. Además, las opciones preestablecidas que acompañan a las preguntas cerradas, pueden ayudar a clarificar la pregunta misma al encuestado, así se evitan problemas de mal interpretación y de respuestas inconsistentes.

Desde la perspectiva impresionista de los diseños cualitativos, se insiste en que con las preguntas cerradas se fuerza a los encuestados a limitar sus apreciaciones dentro de categorías que supuestamente no cubren la cualidad del hecho o proceso sobre el cual se está inquiriendo y en consecuencia, la respuesta tiene menos significado y menos capacidad de aprehensión. Esta distorsionada imagen tiene asidero tanto en el abusivo uso de preguntas cerradas con escasas alternativas -dos o tres categorías- o, en el uso de la *categoría otros* sin permitir que el encuestado señale y consigne la posible respuesta alternativa.

En cuanto a los cuestionarios con preguntas abiertas, la ventaja es su utilidad para *explorar procesos o problemas desconocidos parcial o totalmente* por los investigadores. Este tipo de formato da la posibilidad al encuestado de describir la razón de sus ideas. Algunas veces, en el proceso de construcción de un cuestionario es difícil especificar todas las posibles respuestas a una determinada pregunta. En esta situación una pregunta abierta es un excelente instrumento para lograr cubrir todas las posibilidades de respuesta, para posteriormente codificarlas. Estas preguntas son de gran valor cuando se quiere medir el conocimiento de los encuestados sobre un determinado hecho o personas involucradas en él.

Lo extraño es que la principal desventaja de *las preguntas abiertas* esta relacionada con su ventaja principal, según la cual el encuestado puede extenderse en sus respuestas; pero esto puede dificultar enormemente el proceso de *validación, crítica y codificación de la información* obtenida. Ambas situaciones son relevantes si finalmente se decide que se procesará esa información utilizando un programa informático especializado; y esto es difícil, pues las preguntas abiertas deben cerrarse *tipificando*³¹ las distintas respuestas y después vaciarse éstas en hojas de codificación antes de su digitación en un programa informático.

Cuando no es posible que la respuesta a la pregunta sea completamente cerrada, es frecuente hacer una combinación y proponer varias respuestas, aunque dejando la posibilidad de una respuesta abierta, lo que da origen a cuestionarios mixtos:

Ocupación de la persona que contribuye principalmente al sostenimiento del hogar:

- | | |
|----------------------|----------------|
| 1. Empleado | 4. Comerciante |
| 2. Obrero | 5. Profesional |
| 3. Por cuenta propia | 6. Otro |
- Especificar: _____

- Limitaciones de la técnica de la encuesta social

En el caso particular de los sondeos de opinión, éstos son una foto instantánea, una radiografía de la opinión pública. La calidad de la fotografía depende no sólo de la calidad

³¹ El proceso de *tipificación* consiste en encontrar cuáles son las respuestas más comunes, asignarles una clave y después, con esas claves, codificar las respuestas de todos los cuestionarios.

de la máquina fotográfica sino también de la capacidad creativa y de manejo del fotógrafo- no hay máquinas fotográficas automáticas que produzcan encuestas en serie - El investigador define el campo u objeto, la abertura del lente, la distancia, revela y enmarca la instantánea. Teóricamente sólo registra y mide la opinión. Como estamos hablando de procesos en donde la mano del hombre juega un papel importantísimo, pueden darse fotos movidas, oscuras o fuera de foco o simplemente trucas, ante esta situación cabe preguntarse si esto es producto de un mal manejo técnico o si es el efecto de que los sujetos-objeto de la fotografía, por sus propias características, sean imposibles de *captarlos* o que haya cierta intencionalidad de manipulación con preguntas que tergiversen o induzcan las respuestas. Esto último tiene que ver con la utilización de los datos obtenidos: puede ser la materia prima para hacer un fotomontaje para crear opinión -consensos consumistas o políticos, etc.- o ser una técnica que permite oscultar la opinión, suministrar elementos de juicio a los hacedores de políticas; en síntesis, un estudio de opinión *capta* los acuerdos o desacuerdos de los pensamientos de los demás con relación a hechos sociales, coyunturas, imágenes, etc.

Aquí se afirma, que los sondeos mediante encuestas son como una fotografía instantánea, y eso implica que en la próxima fotografía todo será diferente. Es por eso que se recomienda, especialmente en los estudios de opinión sobre tendencias electorales, que haya continuidad en el tiempo de los sondeos.

También se presentan algunas limitaciones técnicas con las preguntas, es por tal razón que, se recomienda tener en cuenta las características de las que se van a utilizar para medir las dimensiones - variables del estudio.

Cuando se comienza a formular un conjunto de preguntas, a partir de las dimensiones del estudio, se debe usar *un vocabulario apropiado de acorde a la población* y se deben tener

en cuenta algunas recomendaciones, que más adelante detallamos, para evitar sesgos en las respuestas. Por ejemplo, se recomienda evitar *preguntas dobles* -dos preguntas en una- tal como la siguiente:

¿Esta Ud. de acuerdo con la declaración del Gobierno que la política de apertura económica es irreversible y que se debe congelar el gasto público para facilitar el proceso?

Con la pregunta anterior se pueden dar situaciones como las siguientes: una persona puede estar de acuerdo con la política de apertura pero en desacuerdo con la reducción del gasto público. Y, viceversa, estar en desacuerdo con la política de apertura y, de acuerdo con una mayor intervención del Estado reflejado en una mayor gasto fiscal.

Para el analista, el problema es similar: ¿cómo interpretar una respuesta que señale un acuerdo? ¿Este es aplicable a ambos componentes de la pregunta o, sólo a uno de ellos?

Del enunciado de la pregunta citada, surgen otros problemas, sobre los cuales hay que tomar algunas precauciones o, al menos tener conciencia de sus efectos. A pesar del lógico cuidado y esfuerzo del investigador, en el proceso de construcción del cuestionario, al utilizar conceptos cuyas definiciones operacionalizables permitirían una mayor y mejor aprehensión de la cualidad que se está estudiando, se deben evitar conceptos o palabras con *múltiples significados o interpretaciones* que puedan distorsionar los objetivos que se quieren lograr con tal o tales preguntas. Un típico ejemplo de una pregunta ambigua es:

¿Que opina sobre la legitimidad del Estado y la gobernabilidad democrática?

Tal formato de pregunta no sólo es ambiguo en el planteamiento sino también, en lo que realmente se quiere inquirir. Tal como se aprecia, una pregunta deficientemente planteada puede implícitamente abarcar más de un tema, sino también ser ambigua o vaga conceptualmente. La sofisticación conceptual, como el uso de anglicismos o conceptos técnicos, al igual que el uso de modismos regionales o un vocabulario muy académico parecen ser problemas recurrentes en el diseño de un cuestionario.

Esta sugerencia metodológica toca con otro problema que es el que cada cuestionario debe ser construido de acuerdo, no sólo por los objetivos del estudio, sino también teniendo en cuenta las características de quien responde - características demográficas, socio-económicas, etc.

Otra situación de sesgo puede darse cuando la información pasa del investigador al divulgador. Con el auge de los estudios de tendencias electorales a través de los estudios de opinión, aparece también la polémica: ¿Que tan confiables son? ¿Los resultados que arrojan crean o manipulan opinión? ¿Hay manipulación de los medios de comunicación por la forma que difunde la información? Una posible respuesta la da Christopher Hitchens quien al hacer un recuento histórico señala:

"Las encuestas fueron el fruto de una lucha no para descubrir la opinión, sino para manejarla. Fue un arma no para descubrir la opinión, sino para manejarla. Fue un arma en las primeras guerras para frustrar los planes del obrerismo organizado y en la batalla contra el populismo; y más tarde fue el recurso favorito en el arsenal de la sicología de masas." (Hitchens, Christopher, 1998:8).

Hitchens también hace referencia a aquellas preguntas que pueden inducir *respuestas esperadas o de preferencia obligada* que se presentan cuando el interrogador plantea una proposición tajante o que dan pie a elegir un sólo tipo de respuesta.

De igual manera, las preguntas demasiado escuetas o concisas en la formulación del interrogante y amplias o recargadas en los contenidos implícitos o explícitos (sobre la temática), según Hitchen *"pueden, así provocar una reacción nerviosa y suscitar respuestas impensadas"*.

Las preguntas acerca de temas personales o íntimos tales como preferencias políticas, el aborto, uso de anticonceptivos, drogadicción deben ser diseñadas, de tal forma que no necesariamente exijan una respuesta que involucre explícitamente al encuestado y que no provoquen su rechazo. Se sugiere primero, hacer las preguntas en términos que al encuestado se le interrogue acerca de lo piensan **los otros** alrededor del tema (para que no se sienta directamente involucrado), de esta forma, la respuesta obtenida reflejará más la opinión del encuestado que la de los "otros". Metodólogos como Hyman Korman³² afirman que en estos casos la respuesta reflejará realmente la opinión del encuestado más que la de **terceros**. De igual forma, se sugiere utilizar las técnicas escalares de medición de actitudes en lugar de preguntas directas para el tratamiento de este tipo de temáticas o para medir preferencias o niveles de satisfacción.

Formular preguntas de manera que implique aceptar ciertos presupuestos implícitos puede convertirse en un instrumento de manipulación de la opinión, en el sentido de hacer admitir aspectos íntimos que con dificultad se reconocerían abiertamente, como por ejemplo:

Estamos estudiando los cambios recientes en las prácticas anticonceptivas de las universitarias. ¿Cuáles de estos anticonceptivos usa usted?

Anticonceptivo oral.- Dispositivos intrauterinos - Esterilización u otros.

³² Department of Sociology. State University of New York at Stony Brook, New York, 1984.

Este tipo de preguntas coloca al interrogado, en una situación comprometedoras ya que se da por sentado que la persona tiene una práctica anticonceptiva. Asimismo, vincular instituciones o personalidades a temáticas, como una forma de respaldar socialmente o como evidencia comprobada puede afectar las respuestas. Lo anterior se constituye en una precaución metodológica a tener en cuenta, si se quieren evitar sesgos en las mismas.

Otro problema con las respuestas puede suceder cuando se confronta al encuestado con una serie de enunciados o afirmaciones sobre diferentes temas –técnicas escalares de medición de actitudes -y dado el diseño del cuestionario se le ofrezcan idénticas categorías de respuesta, por ejemplo:

Muy de acuerdo – Acuerdo - Indiferente – Desacuerdo - Muy en desacuerdo

El encuestado puede estar de acuerdo con los primeros enunciados y luego por efecto del cansancio o reticencia a la encuesta, declarar reiteradamente estar *de acuerdo* con los planteamientos o afirmaciones siguientes. Este fenómeno es lo que se llama *desviación de convencimiento*: Se tiende a responder **sí** antes que **no, de acuerdo** antes que **en desacuerdo**.

Para evitar las respuestas *rutinarias*, se sugiere que los enunciados o ítems relacionados con el fenómeno que se va a medir sean redactados, en forma alternada, de tal forma, que el encuestado pueda estar de acuerdo con el primer enunciado y no necesariamente con el segundo y así sucesivamente; de esta forma se obliga a poner atención y a responder cuidadosamente. También, el continuo de respuesta que representa una



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

escala ordinal, de un mayor valor a un menor valor – de totalmente de acuerdo a totalmente en desacuerdo- deberá ser presentado *alternadamente*.

Hay temas, como la inseguridad ciudadana, el terrorismo, el secuestro, que determina un tipo de respuesta emotiva, no pensada; opinión posiblemente inducida por los medios de comunicación. La sutileza en su formulación y la ubicación dentro del formato del cuestionario son recomendaciones a tener en cuenta. El hecho de ser presentadas al comienzo del cuestionario puede influir en las respuestas posteriores, posiblemente a causa del esfuerzo del encuestado de aparecer consistente, con sus afirmaciones primeras o con respuestas socialmente aceptables.

Semejante a lo anterior, es el fenómeno que se presenta en las encuestas de tendencias o vaticinios electorales y de imagen. Al colocar preguntas directas sobre la imagen de una persona o de preferencias políticas, al comienzo del cuestionario, se motiva a que la tendencia a responder *no sabe o no opina* sea mayor que si estas preguntas estuviesen ubicadas en un orden posterior.

Lo anterior, puede ser explicado por la tendencia del encuestado a responder de manera tal que no se comprometa o en una forma socialmente aceptable. Esta situación aparentemente, es superada si este tipo de preguntas se hace en un momento en el cual el encuestado se siente más en confianza y motivado con los temas de la encuesta.

Y ya que nos estamos refiriendo a las encuestas de tendencias electorales, digamos que en este campo hay que ser cuidadoso. Una de las principales características de este tipo de sondeo de opinión es que son dinámicas, porque permanentemente están retratando una imagen de algo tan voluble como la opinión pública.

Las cifras de las tendencias electorales que barajan los medios de comunicación, son justamente eso, tendencias no inclinaciones electorales definitivas y de todas maneras



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

cada sondeo refleja más una imagen que una intención de voto, es decir, el encuestado a meses de la elección, cuando responde el cuestionario no está pensando en su intención electoral definitiva.

Los metodólogos Webb y Hatry en su texto *Application of Citizens Survey to Local Governments*, (Webb, K., Hatry, H., 1973: 33-39), dan una serie de recomendaciones acerca del uso de los estudios de opinión:

- *Cuidado con los estudios de opinión sobre temas complejos que la gran mayoría de los ciudadanos no conoce o no tienen mayor información.*
- *Cuidado con las respuestas que reflejan opiniones o apreciaciones prematuras o tempranas producto de un momento coyuntural de crisis societal o personal; son opiniones que hay que tener en cuenta pero dentro de ese contexto al igual que aquellas que reflejan situaciones a más largo plazo.*
- *Cuidado con los estudios de opinión sobre temáticas que pueden interferir o inclinar las preferencias en una contienda electoral próxima.*
- *Cuidado con el enunciado conceptual de una pregunta. Lo que se diga o se deje de decir puede prestarse para malentendidos.*
- *Cuidado con resultados no representativos, producto de inadecuados procedimientos de selección, diseño o contenido de la encuesta. Ejemplo de esto puede ser el rechazo generalizado a responder una determinada pregunta o cuestionario. Si esto sucede se puede deber a una inadecuada posición de la pregunta, una población muestral sesgada, etc., lo anterior daría motivos a considerar inválidos los resultados obtenidos.*

Dedicar atención a estas recomendaciones básicas, permite aumentar la probabilidad que el encuestado interprete las preguntas adecuadamente permitiendo respuestas confiables y validas. Los estudios de opinión ayudan a averiguar que piensa la gente, pero su mayor importancia a largo término, esta en el hecho que cumple una función importante en el proceso de comunicación del ciudadano con las instituciones.

- La composición y orden de las preguntas de un cuestionario

La redacción, el ordenamiento y disposición de las preguntas son pasos básicos en el diseño de un cuestionario. No debe olvidarse que las preguntas son estímulos verbales que deben despertar el interés sobre el tema y por tanto deben ir creando una atmósfera de confianza y motivación a contestar a las demás preguntas. Por eso es tan importante el ordenamiento de las preguntas. Elizabeth Noelle refiriéndose a lo que ella llama "el cuestionario con buenos modales" dice lo siguiente:

"Cuestionarios que pueden tener buenos o malos modales, pueden ser cortas o atento, o tosco o egoísta" [...] por ejemplo, la elección de las preguntas de contacto al comienzo, o la colocación de una pregunta interesante o alegre al fin de la entrevista [...] Pero, sin darnos cuenta, las preguntas de buenos modales en la entrevista se convierten en condiciones cuyo cumplimiento determina inmediatamente la calidad de los resultados", (Noelle, Elizabeth, 1970: 134).

El orden en la presentación de las preguntas, es otro criterio metodológico que hay que tener en cuenta, ya que puede influir en la confiabilidad y validez de las respuestas. Por ejemplo, al ordenar las preguntas dentro del formato del cuestionario se debe tener en cuenta, la influencia de las preguntas previas en las respuestas siguientes -*contaminación de las preguntas*-.

Según se coloque una pregunta al principio o al final del cuestionario, puede provocar respuestas diferentes. A medida que se va desarrollando el interrogatorio el encuestado reflexiona, se familiariza, se forma una opinión sobre el tema, todo lo cual incidir sobre el contenido de su respuesta. Cuestión diferente es que se le formulen preguntas inesperadas en relación con su posición dentro del cuestionario, dicha posición debe ser cuidadosamente escogida.

Ghigliione refiriéndose a lo que él denomina *el efecto de halo* afirma que ,este se manifiesta "*cuando se solicita primero un juicio global: ciertas personas se sentirían incoherentes si, al haber expresado de entrada un juicio favorable, manifestaran después críticas o reservas sobre aspectos particulares.*" De igual forma, previene acerca de los problemas que conlleva "*formular múltiples preguntas sobre el mismo tema o sobre temas próximos; tal es el caso cuando se pretende construir una escala de actitud o se desea explorar los diferentes aspectos de un tema y analizar las relaciones entre las respuestas. Es necesario, entonces combinar enunciados favorables y desfavorables.*" (Ghiglioni, 1989:134). A veces es recomendable la alternación de temas en las preguntas para lograr la confianza y motivación del encuestado, de tal forma que se logren respuestas de mayor profundidad.

Para prevenir *la contaminación* de las preguntas se pueden seguir estos criterios:

- Dispersar las preguntas en el conjunto del cuestionario, en forma de alejar una de otra aquellas que arriesgan contagiarse.
- Se puede disponer las preguntas en forma de "*embudo*", es decir, colocando primero aquellas preguntas más generales y después progresivamente las particulares y especiales, evitándose así las sugerencias de unas a otras.

4. Conclusión

A lo largo de la Capítulo se ha intentado que el estudiante se familiarice con la medición en la investigación social que utiliza preferentemente los diseños cuantitativos y los instrumentos de recolección de información primaria. Muchos diseños de investigación requieren la recolección sistemática de información sobre poblaciones totales o de sectores representativos (muestras poblacionales). Una vez seleccionado el diseño de investigación de acuerdo con el problema de estudio y sus hipótesis, el paso siguiente



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

será recolectar la información pertinente. Se presentaron algunas de las técnicas de recolección de información más utilizadas en los diseños cuantitativos que facilitan el manejo de la información cuantificable.

- Ejercicio de autoevaluación N° 3

Proceso de elaboración de un cuestionario mediante la técnica de la encuesta social.

Diseñe un cuestionario acerca de una temática conocida por Ud. que combine varios tipos de preguntas.

Nota metodológica

El cuestionario es un formulario que contiene preguntas o variables operacionalizadas del estudio. En términos operacionales, se sugiere diseñar una matriz de análisis con las dimensiones-variables que arrancan del problema de investigación. Los conceptos - dimensiones y variables mediante los cuales se ha planteado su problema de investigación y los objetivos de ella, son la base de las preguntas de un cuestionario. Pero para poder traspasar de los conceptos e hipótesis a las variables y de éstas a las preguntas se necesita un guía – *el indicado*³³ - de lo que debe observarse. El esquema es el siguiente:

Dimensiones ⇒-Conceptos ⇒- Variables - Indicadores - Preguntas

Proceda a formular las preguntas referidas a las dimensiones – variables que usted haya seleccionado – no más de seis - teniendo en cuenta las sugerencias metodológicas, acerca de cómo formular una pregunta, planteadas en esta Capítulo de tal modo que la información recogida en el interrogatorio sea confiable y válida.

³³ El indicador hace mensurable observables a los conceptos - variables no directamente medibles y observables. Este proceso de convertir una *dimensión* en *indicadores* se le denomina operacionalización.



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

ceo@catios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

CAPÍTULO 5. SISTEMATIZACIÓN DE LA INFORMACIÓN

1. Introducción

El presente Capítulo contiene la descripción breve de la secuencia de procedimientos utilizados para *preparar* una información para ser sistematizada. El procesamiento de la información requiere no solo de un proceso de planeación sino también de un plan adecuado de manipulación de los datos lo que determinará unos resultados que realmente sirvan para los objetivos del estudio. Igualmente, se muestran los procedimientos para configurar la estructura de un archivo de trabajo, introducir los datos, guardarlos, transformarlos y leerlos.

El procesamiento de los datos es una necesidad muy usual en diversas áreas: gubernamental, de investigación, académica, etc. Aunque los campos de aplicación sean diferentes, muchos de estos procesos son comunes. Dada esta estandarización de la demanda de necesidades se han desarrollado programas especializados de manejo y gestión de datos. Existe una gran variedad de estos programas que difieren entre sí, por ejemplo, en los aspectos de capacidad –número de datos y variables-, variedad de cálculos que pueden realizar y facilidad de su uso. Uno de estos programas es el SPSS (Statistical Package for the Social Sciences)³⁴, diseñado en forma modular, cada uno con menús desplegables y cuadros de diálogo que permiten acceder a las distintas funciones que ofrece el programa. Este programa facilita en forma sencilla, dos de las tareas básicas del proceso de investigación social aplicada, es decir, la entrada y grabación de los datos obtenidos y el análisis estadístico de los mismos.

2. Objetivos

³⁴ SPSS es marca registrada de SPSS. Inc.

Ante un conjunto de datos a analizar, surge la necesidad de elegir el tratamiento de sistematización, es decir, la elección de una metodología apropiada. Para una correcta escogencia es necesario conocer los medios que se dispone.

Este Capítulo tiene como objetivo: Familiarizar a los estudiantes en el proceso de sistematización. Es decir introducir a los estudiantes en el tratamiento de los datos. Por lo general, supone un tratamiento informático de los datos, lo cual implica la digitación o grabación, y el almacenamiento en archivos de trabajo para su posterior manipulación y análisis. En este Capítulo, se abordarán los aspectos operativos relativos al diseño de archivos de trabajo mediante los módulos del programa SPSS.

3. Desarrollo

Desde el momento que se comienza a estudiar el planteamiento de un problema de investigación hay que planear dentro del diseño operacional, el proceso de sistematización de los datos.

3.1. Sistematización de la información

El procesamiento de la información obtenida en el trabajo de campo requiere de un plan adecuado de manipulación de los datos y para tal efecto se deben realizar las siguientes actividades:

3.1.1. Plan de análisis de los datos: De acuerdo a los objetivos del estudio se debe previamente diseñar un plan adecuado de manipulación de los datos.

3.1.2. El proceso de codificación de la información obtenida: A partir de la información obtenida y una vez diseñado el cuestionario con las preguntas, se debe realizar la

llamada prueba piloto para validar el instrumento de recolección de información. Probado el instrumento se debe proceder al proceso de recolección de información. Una vez validados los datos recogidos, se deben clasificar las respuestas individuales obtenidas. Esta etapa se denomina codificación y es de gran importancia ya que la traducción de todas las respuestas a códigos exige no sólo de un conocimiento del tema sino también de los objetivos del estudio.

El proceso de codificación consiste en la asignación de un código numérico a las diferentes categorías de respuestas a cada pregunta del cuestionario. El instrumento para transformar las respuestas en datos estadísticos es el libro de codificación.

3.1.3. Diseño del libro de codificación: Se recomienda diseñarlo de tal forma que aquellos que realicen el trabajo de codificación tengan una guía precisa de cómo asignar un código a cada uno de los valores de las categorías de respuesta de cada pregunta (variable). El libro permite traducir cada categoría de respuesta a un número preestablecido (código numérico), el cual es vertido en la hoja de codificación.

Nota metodológica

Veamos algunas recomendaciones que hay que tener en cuenta en el proceso de codificación:

- Al diseñar los códigos se debe tener en cuenta los objetivos del estudio y la variedad de respuestas que encontremos en los cuestionarios.
- Recuerde siempre que por cada pregunta sólo se debe aceptar una sola respuesta ya que la mayoría de los programas estadísticos, a no ser que se acepten multirespuestas.
- Crear tantos códigos como posibles categorías de respuesta existan. No olvidar que un número superior a 10 categorías dificulta enormemente la construcción de

tablas y figuras. En el caso, por ejemplo de clasificaciones de oficios, en donde es posible encontrar innumerables categorías, se sugiere crear tantos registros de preguntas como sea necesario, pero posteriormente recodificar o agrupar en no más de 10 categorías.

- Se recuerda que en los datos nominales, el valor numérico del código asignado no significa una jerarquía ni orden.

Ejemplificación de un libro de codificación

Variables	Categorías	Código
Sexo	Masculino	1
	Femenino	2
Estado civil	Soltero	1
	Casado	2
	Viudo	3
	Separado	4
	Unión libre	5

Edad en grupos quinquenales

	15	19	1
	20	24	2
	25	29	3
	30	34	4
	35	39	5
	40	44	6
	45	y más	7
Nº de años de Vinculación	1	9	1
	10	19	2
	20	29	3
	30	39	4
	40	49	5
	50	y más	6

- **Transferencia de datos a una hoja de codificación:** Una hoja de codificación es una matriz formada por fila y columnas; una fila (horizontal) representa cada uno de los



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

registros –cuestionario- de la encuesta; y las columnas (verticales) representan el valor de cada variable.

Ejemplificación

La variable 1 *Sexo* es un campo que ocupa una columna en la hoja de codificación. Supongamos que la variable *sexo* ocupa la columna N° 1. En esa celdilla anotaremos el valor que tome la variable *sexo* en cada uno de los sujetos del estudio. El dato de la variable N° 2: *Estado civil* (campo que ocupa un espacio –un dígito) lo anotaremos en la columna 2 y así sucesivamente.

3.1.4. Generación de archivos de trabajo: Superadas las etapas de crítica y codificación hay que realizar la digitación de la información, contenida en las hojas de codificación, en la base de datos.

El proceso en sí comprende las siguientes etapas:

Grabación, depuración, edición y tratamiento de los datos

En este Módulo se utilizará un programa de manejo y gestión de datos: El SPSS, que es un programa de análisis de información general y procesamiento estadístico. Sus siglas *SPSS* significan **Statistical Package for the Social Sciences**, originalmente concebido como un programa para este tipo de análisis, especialmente rico en métodos no paramétricos y paramétricos.

El SPSS fue concebido como un sistema de análisis de información. Con él, es posible efectuar las siguientes operaciones:

- Crear y editar archivos de datos.
- Recodificar variables y calcular nuevas variables.
- Grabar datos hacia archivos de datos recuperables o de formato específico.

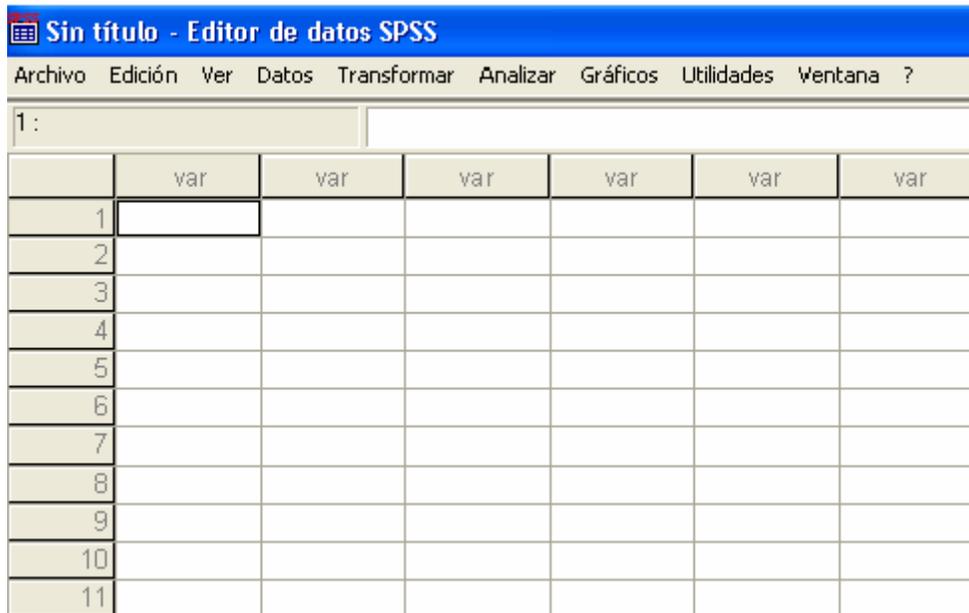
- Tabular variables y combinaciones de variables.
- Edición y transformación de datos
- Efectuar una serie de análisis estadísticos.

Como programa de gestión y manejo de información el SPSS permite entrar datos interactivamente desde un archivo externo, recodificar variables y calcular nuevas variables, grabar y almacenar datos hacia archivos de datos recuperables, transferir archivos de datos entre computadores, tabular variables y combinaciones de variables, efectuar una serie de análisis estadísticos, producir reportes, etc.; pero el SPSS es sobretodo, un programa de procesamiento estadístico y sus méritos fundamentales radican en la amplitud y el rigor de las pruebas estadísticas que contienen.

3.1.4.1. Manejo operativo del SPSS

SPSS para Windows consta de **ventanas, menús, herramientas, cuadros de diálogos** que permiten comunicación con el usuario. Aquí se incluye una breve descripción de estos conceptos para unificar el lenguaje.

Para iniciar el programa *haga doble clic con el botón izquierdo del ratón sobre el icono del SPSS*, aparecerá la ventana del Editor de datos SPSS y la ventana del Visor SPSS.



	var	var	var	var	var	var
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						

A través de otras ventanas, de carácter temporal, SPSS interactúa con el usuario para brindar ayuda sobre una acción específica (*ventanas de ayuda*) o para dar mensajes de error o advertencias (*ventanas de mensajes*) o para completar información necesaria para desarrollar los procesos necesarios para el análisis de los datos. Estas últimas ventanas se denominan *cuadros de diálogo*.

Menú principal del SPSS

La pantalla principal del SPSS presenta una barra de menús con 9 menús desplegables tipo cascada.

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventanas

Descripción de los menús:

Archivo: Utilice este menú para crear un nuevo archivo de SPSS, abrir un archivo ya existente o leer archivos de bases de datos ya creados.

Edición: Utilice este menú para modificar o copiar textos de las ventanas de resultados.

Ver: permite visualizar la barra de estado y la barra de herramientas.

Datos: Este menú permite realizar cambios globales en los archivos de datos de SPSS, tales como la fusión o división de archivos, la transformación de variables y casos o la creación de subgrupos de casos para su análisis.

Transformar: Utilícelo para realizar cambios en variables seleccionadas en el archivo de datos y para calcular nuevas variables basadas en los valores de las ya existentes.

Analizar: Utilice este menú para seleccionar los diferentes procedimientos estadísticos que quiere utilizar, como tablas de contingencia, análisis de varianza, correlación y regresión lineal, etc.

Gráficos: Este menú permite crear gráficas de barras, gráficas de sectores, histogramas, gráficos de dispersión y otras gráficas de alta resolución y a todo color. Algunos procedimientos estadísticos también generan gráficas.

Utilidad: Utilice este menú para visualizar información acerca del contenido de los archivos de datos y sus variables, o abrir un índice de los comandos de SPSS.

Ventana: Utilice este menú para disponer, seleccionar y controlar los atributos de las diferentes ventanas de SPSS.

? (Ayuda): Este menú abre una ventana de ayuda estándar de Microsoft que contiene información acerca de la utilización de las diversas operaciones del SPSS.

- Procedimiento para seleccionar las opciones del menú

Llevar el puntero del ratón a la barra de menús y señalar uno de ellos, y al darle clic se desenrollan hacia abajo mostrando las opciones o submenú de trabajo. En la ventana se desplegará el *menú* seleccionado, mostrando los comandos disponibles para esa opción.

En una segunda selección el despliegue puede ser a su vez un *sub. - menú* dentro del cual se selecciona una opción de la misma forma. Se puede seleccionar a través de una cascada haciendo clic sobre la palabra deseada de la barra de menús y sin liberar el botón del ratón, moverlo hasta que su puntero se ubique sobre los artículos de los sub. - menús uno tras otro.

- Botones de comando

Existe una columna de botones de comando en el lateral derecho del cuadro de diálogo. Frecuentemente se les denomina **botones de comando** porque piden al sistema que realice alguna función de forma inmediata, como ejecutar un procedimiento u obtener ayuda. Existen cinco botones de comando estándares en el cuadro de diálogo principal:

Aceptar: Ejecuta el procedimiento seleccionado. Después de seleccionar las variables que desee y de elegir cualquier especificación adicional, de un clic en Aceptar para ejecutar el procedimiento. Este botón también cierra el cuadro de diálogo.

Pegar: Genera una sintaxis de comandos a partir de las selecciones del cuadro de diálogo y pegar la sintaxis en una ventana de sintaxis.

Restablecer: Anula la selección de cualquier variable de la(s) lista(s) de variables seleccionadas y restablecer todas las especificaciones del cuadro de diálogo principal y cualquier cuadro de diálogo al modo predeterminado.

Cancelar: Cancela cualquier cambio en los valores del cuadro de diálogo desde la última vez que se abrió y cierra el cuadro de diálogo.

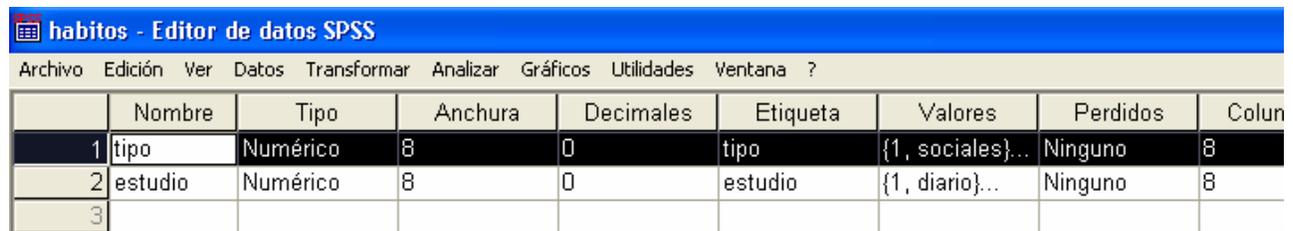
Ayuda: Le lleva a la ventana de Ayuda estándar del programa.

- Entrada de datos

La matriz del editor de datos es un rectángulo formado por filas y columnas: Cada fila representa un caso o una observación. Por ejemplo cada individuo que responde a un cuestionario es un caso. Cada columna representa una variable o característica que está siendo calculada. Por ejemplo, cada elemento de un cuestionario es una variable. Cada celda contiene un sólo valor de una variable para un caso. La celda es la intersección del caso y la variable. Las celdas contienen sólo valores de datos. Las dimensiones del archivo de datos se determinan por el número de casos y variables. Puede introducir datos en cualquier celda. Para variables numéricas, las celdas en blanco se convertirán en valores perdidos por el sistema.

- Configuración de la matriz de variables

Como opción predeterminada, SPSS da por sentado que todas las variables nuevas son numéricas. Para cambiar el tipo de variable, de un doble clic en la variable *Tipo* en el editor de datos SPSS. Esto abre el cuadro de diálogo Vista de variables, como se muestra en la siguiente figura.



	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columna
1	tipo	Numérico	8	0	tipo	{1, sociales}...	Ninguno	8
2	estudio	Numérico	8	0	estudio	{1, diario}...	Ninguno	8
3								

En Vista de variables, Ud. puede nombrar las variables, definir el tipo de variables, el ancho de los datos en la columna, si tiene o no decimales los datos, la etiqueta o descripción de la variable, los valores o categorías de respuesta de la variable, los valores perdidos y el número de dígitos que ocupa el nombre de la variable.

- Nombres de las variables

El nombre predeterminado para las variables nuevas es el prefijo **var.** y un número secuencial de cinco dígitos (**var00001, var00002 etc.**). Para cambiar el nombre de la variable, sólo tiene que introducir el nombre nuevo de la variable.

- Digitación de los datos

Una vez definida la estructura de la matriz de variables en el editor de datos comienza la digitación de los datos contenido en la hoja de codificación. Para ello de un clic en la solapa *Vista de datos* y se comienza el procedimiento de digitación por columna. Una vez digitado todos los datos de la primera columna –variable *Tipo* – se continúa con la siguiente columna –variable *Estudio*- y así sucesivamente.

- Grabación de los datos digitados

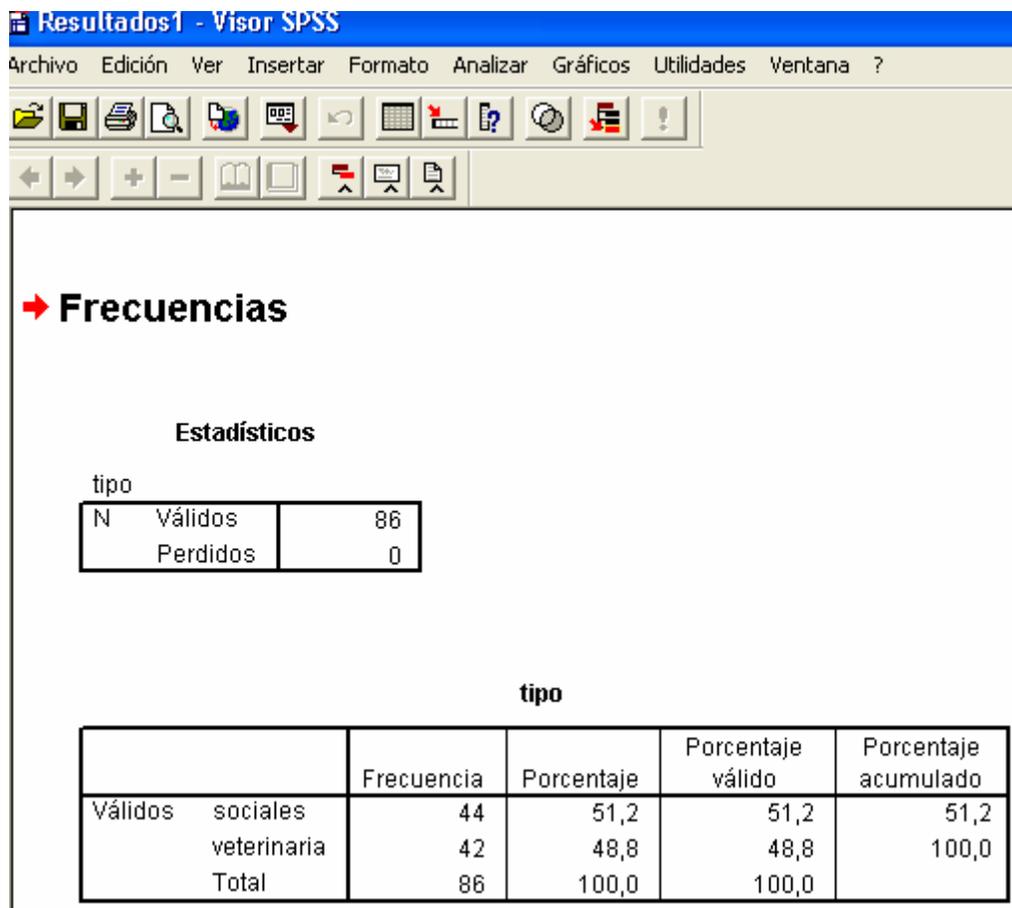
Para grabar los datos hay que situarse en el Editor de datos y elegir Archivo ⇒ Guardar. Si es un archivo nuevo hay que escribir el nombre de él en la celdilla al lado de Nombre del archivo y seleccionar en *Guardar como archivos de tipo: SPSS* (habitos.sav).

- Lectura de datos

Para leer datos guardados anteriormente elegir Archivo ⇒⇒ Abrir ⇒⇒ Datos. Seleccione en el cuadro de diálogo con todos los archivos almacenados.

- El visor de resultados

Este cuadro contiene los resultados en forma de texto de su sesión de trabajo con el SPSS. Esto incluye cualquier resultado estadístico no gráfico, como una tabla de contingencia o una matriz de correlaciones. Puede editar y guardar los resultados en archivos de texto para un uso posterior. Por ejemplo, se realizó un procedimiento de distribución de frecuencias con una variable *tipo* del archivo *hábitos.sav*. En el visor se puede ver que el 51 por ciento de los estudiantes son de la Facultad de Ciencias Sociales y el 48.8 son de Veterinaria.



Resultados1 - Visor SPSS

Archivo Edición Ver Insertar Formato Analizar Gráficos Utilidades Ventana ?

→ **Frecuencias**

Estadísticos

tipo

N	Válidos	86
	Perdidos	0

tipo

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos sociales	44	51,2	51,2	51,2
veterinaria	42	48,8	48,8	100,0
Total	86	100,0	100,0	



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

ceo@catios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

3.1.5. Procesamiento de los datos obtenidos con el propósito de generar informes:

Una vez conformado y depurado el archivo de trabajo, desde el punto de vista cuantitativo, el analista debe revisar su plan de análisis para asegurar cual es la estrategia de análisis más adecuada que necesita para resolver el problema de investigación.

CAPÍTULO 6. ANÁLISIS E INTERPRETACIÓN DE DATOS

1. Introducción

El analista social debe saber que tipo de estadísticos son más pertinentes y cómo analizar o leer los resultados. La importancia de lo anterior se basa en la idea que el analista no debe ciegamente usar unos estadísticos, es decir, hay que evitar que sea incapaz de juzgar su pertinencia en relación con el problema de estudio, y menos leer y evaluar la calidad de los resultados. Generalmente, este analista *ciego* considera que la objetividad de los resultados está avalados por el carácter que toma todo resultado estadístico cuando es producto casi mágico de un proceso de sistematización, lo cual conduce a muchos errores. Para facilitar ese objetivo, se desarrollan una serie de ejercicios prácticos basados en submuestras pequeñas acompañados de los procedimientos operativos a través del programa SPSS. Ello permite al estudiante seguir esos ejercicios, realizando personalmente los procedimientos de decisión y manejo de los datos. Igualmente se ha elegido una notación de instrucciones que permite operar el programa SPSS que se emplea en esta Capítulo para analizar los distintos ejercicios prácticos presentados.

2. Objetivos

Familiarizarse con algunos instrumentos estadísticos de análisis de la información:

Aplicar técnicas no paramétricas:

Descripción de la información: aplicaciones de la estadística descriptiva

Utilización de estadísticos para variables nominales y ordinales

Asociación de variables nominales y ordinales

Aplicar técnicas paramétricas:

Regresión y correlación

Técnicas de análisis multivariado

Hacer inferencia estadística y prueba de hipótesis

3. Desarrollo

Una vez planteado el problema que se va a investigar y se ha definido el marco de referencia teórico, conceptual y metodológico que guiará el proceso de investigación, es necesario considerar cómo se analizará la información obtenida.

En los estudios sociales se necesita frecuentemente analizar el comportamiento de un conjunto de datos con el objeto de determinar hechos o características que son de interés. Gran parte del trabajo consiste básicamente en analizar una información que se nos presenta, en primera instancia, en forma descriptiva. El término análisis aquí lo vamos a entender, en una doble perspectiva, una como un proceso de búsqueda de elementos o aspectos o configuraciones que caractericen y contextualicen los procesos o hechos estudiados. Y otra, como *una simple lectura* de unos resultados obtenidos. Lo anterior está relacionado con el hecho que el análisis de los datos depende de una serie de elementos, entre los cuales, los más importantes a considerar son:

- Los objetivos del estudio
- La hipótesis de investigación
- El nivel de medición de las variables

Otra consideración a tener en cuenta, cuando se trabaja con un cúmulo de información asociada a una población, son los interrogantes que surgen con relación al comportamiento de los datos, especialmente en lo que respecta a la distribución y asociación de éstos. Briones señala al respecto:

De manera normal, la primera tarea del análisis cuantitativo consiste en obtener una información básica sobre las variables contenidas en el estudio, como es el caso cuando la información se ha recogido con un cuestionario estructurado con un número más bien elevado de preguntas.

Según la naturaleza de las variables, en cuanto a su nivel de medición, esta primera información proporcionará distribuciones de frecuencia, medidas de tendencia central (medias aritméticas, medianas), medidas de dispersión y otras. Todas ellas - o las que el investigador considere más importante- proporcionarán una descripción global del colectivo estudiado como también permitirá orientar el análisis subsiguiente sobre la base de las situaciones más relevantes que hayan podido generar... (Briones, Guillermo, 1988:23)

3.1. Análisis descriptivo de los datos

La primera tarea del analista es, a partir de una información confiable y válida, hacer una descripción de las características de las unidades de análisis, una segunda tarea, si es el caso, hace referencia a la comparación de sub-grupos en relación con una o más características que son importantes en los objetivos formulados por el investigador. Esta secuencia de tareas en el campo del análisis de los datos tiene como objetivo encontrar:

- Aspectos que caractericen, en términos cuantitativos, a los colectivos o procesos, en los cuales se han definido, problemas de investigación –descripción de variables-.
- Clasificaciones, tipologías, diferencias o comparaciones entre ellos.
- Relaciones entre variables sociológicas.

- Descripción de variables

La distribución de frecuencias

Un concepto básico del análisis descriptivo es la distribución de frecuencias que consiste en "el ordenamiento o clasificación de los valores observados en una variable de acuerdo con su magnitud numérica. Permite identificar al investigador la forma como ciertos puntos o características están distribuidos".

La distribución de frecuencias consiste en una tabla que organiza los datos en clases, es decir, en grupos de valores que describen una característica de los datos y tiene la siguiente estructura:

- Categorías de la variable
- Frecuencias: número de casos en cada categoría
- Frecuencias relativas: porcentajes de casos en cada categoría
- Frecuencias acumuladas: porcentaje de casos que se van acumulando en cada categoría desde la más baja hasta la más alta.

Al analizar la distribución de los valores de una variable, el investigador está interesado en:

- Hacer una lectura global de los datos y analizar su distribución: Su frecuencia absoluta o relativa; sus medidas de tendencia central, etc.
- ¿Cómo varían las categorías dentro de la variable?
- Buscar una explicación de la variabilidad de la característica, propiedad o variable.
- ¿Qué factores están asociados a esta variabilidad?

- Procedimiento SPSS

El submenú <**Frecuencias**> es uno de los más utilizados del programa SPSS dado que no sólo sirve para el cálculo de estadísticas, sino también para **visualizar** y **depurar datos** (eliminación de códigos inválidos, modificación de datos mal digitados).

El submenú <**Frecuencias**> elabora una tabla de frecuencias en donde indica cada código que aparece como respuesta, el número de veces que aparece el porcentaje que presentan esas apariciones con respecto al total de casos.

Uno de los productos en **el visor de resultados** de <Frecuencias> es *una tabla distribución de frecuencias* que muestra *las frecuencias relativas* que son los porcentajes de casos en cada categoría. Las distribuciones de frecuencias relativas pueden presentarse en forma de histogramas o gráficas de otro tipo, tales como polígonos de frecuencias³⁵.

- Ejemplificación

La inseguridad es un miedo urbano que permeó todas las esferas. La sospecha nació como mecanismo de defensa y ya muchos ni siquiera confían en el vecino. Los más intrépidos dicen no sentir ese temor. Para los expertos es claro que las políticas represivas solo brindan sensación de seguridad momentánea, que lo que hay que atacar es el fondo del problema. Un sondeo realizado recientemente, mostró que un 46.4 % de los habitantes de la capital antioqueña se sienten inseguros, mientras un 53.6 % afirmó sentirse seguros. Entre los delitos que más desvelan a los medellinenses están los robos, el terrorismo, el secuestro, las violaciones y los homicidios. El sondeo evidenció que un 35.7% de los interrogados, se sienten inseguros por el fenómeno de la indigencia. Las

³⁵ Los polígonos de frecuencias relacionan las puntuaciones con sus respectivas frecuencias. Se utilizan con variables de tipo intervalos.

personas encuestadas ven como una solución el aumento de pie de fuerza, pero también la del fomento de programas de cultura ciudadana.

Ficha técnica

El sondeo de opinión entrevistó, vía telefónica a 322 personas, mayores de 16 años, de diferentes estratos económicos y sexo. Un 46.4% de los encuestados fueron hombres y un 53.6% mujeres. De estos un 32.1% pertenece a los estratos 1 y 2; el 35.7% al 3; el 21.4% al 4 y el 1^o.7% a los estratos 5 y 6. El sondeo tiene un margen de error de 4%.

- Procedimiento SPSS

- En la barra principal de Menús del SPSS, seleccione el menú **<Archivos>**
- Active la opción **<Abrir>** y elija el submenú **<Datos >**
 - Seleccionar archivo **temor ciudadano.sav** en el CD-ROM
- En la barra principal de Menús, seleccione el menú: **< Analizar> »»» <Estadísticos descriptivos> -»»» <Frecuencias³⁶>**
- En el cuadro de diálogo Frecuencias, se selecciona la variable a describir y se transfiere a la sección Variables
- Seleccione simultáneamente las variables “robos” y “secuestr”

En el Visor de resultados aparece:

Tabla de frecuencia

³⁶ Con la opción **<Frecuencias>** aparece como resultado en el Visor una **Tabla de Frecuencias³⁶** en donde se lista cada categoría o código que aparece como respuesta, el número de veces que aparece y el porcentaje que representa esas apariciones con respecto al total de casos.

robos

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	si	172	61,0	61,0	61,0
	no	110	39,0	39,0	100,0
	Total	282	100,0	100,0	

secuestr

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	si	142	50,4	50,4	50,4
	no	140	49,6	49,6	100,0
	Total	282	100,0	100,0	

- Análisis de los datos

De acuerdo a la información del sondeo de opinión sobre *la Inseguridad ciudadana*, se puede apreciar que los robos – en cualquiera de sus manifestaciones- ocupan el primer lugar (61%) en las percepciones de los medellinenses; el secuestro le sigue en orden de importancia entre estos habitantes (50.4%).

3.1.1. Análisis estadístico de los datos

Una segunda aproximación al análisis de los datos, se centra en el uso de estadísticos descriptivos. Cuando se describe un conjunto de variables se trata de sintetizar sus características en unos índices numéricos que representen adecuadamente a ese conjunto de características. En este texto, básicamente las técnicas que se expondrán serán las medidas de tendencia central, valores percentiles³⁷, dispersión y distribución.

Para lograrlo el analista que ha sometido la información a un proceso de cuantificación debe hacerse tres preguntas básicas:

1. *¿Cómo se presentan los datos? ¿Cuál es la ubicación o localización de los datos?*

Lo que se pretende es describir ciertas características de un conjunto de datos. El procedimiento más utilizado, para dar respuesta a este interrogante, son las *medidas de tendencia central*, que permiten resumir el comportamiento de un conjunto de datos alrededor de un punto prefijado tal como: La moda; la media; la mediana

2. *¿Los datos respecto a sus puntos de agrupación y localización están concentrados o dispersos?*

Para determinar la concentración de los datos o su dispersión alrededor de la media, en una distribución, se usa frecuentemente los conceptos de:

- Rango de variación

³⁷ Valores percentiles. Los valores de una variable cuantitativa que dividen los datos ordenados en grupos, de forma que un porcentaje de los casos se encuentre por encima y otro porcentaje se encuentre por debajo.



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

- Valores percentiles
- Desviación estándar o típica

3. *¿Existe relación en el comportamiento de las variables?*

Para establecer la relación, es decir, cuánto de una variable es explicada por las demás, cuyas observaciones han sido registradas en una escala de intervalo o de razón – continua-, se utiliza:

- Coeficientes de asociación entre variables cualitativas
- Coeficientes de correlación entre variables cuantitativas
- Los modelos de regresión

- Nota metodológica

Un criterio básico que hay que tener en cuenta al seleccionar la prueba estadística es el nivel de medición en que fueron medidos los datos³⁸.

La dirección del análisis o la búsqueda de contextualizantes se hace dentro de ciertas *orientaciones* que son señaladas generalmente por los objetivos del estudio.

En el análisis de los datos, estamos en el campo de las técnicas, en cambio en la interpretación estamos en el campo de la conceptualización teórica.

³⁸ El tipo de escala de medición empleado para obtener los valores de una variable es determinante en la selección de los estadísticos de análisis de los datos.

3.1.1.1. Estadísticos para medir la localización de un conjunto de datos

Una tarea básica del análisis de los datos consiste en obtener una descripción global de las características de los sujetos o procesos – objetos del estudio-. Esta primera información mostrará cómo se distribuyen los valores de la variable. Frecuentemente se necesita resumir descriptivamente un conjunto de valores numéricos para lo cual es apropiado recurrir a medidas de tendencia central. Todas estas estadísticas darán una descripción global de la población estudiada y permitirán orientar el análisis. Existen 3 medidas que pueden ser asociadas a este concepto y ellas son: La Moda, la Mediana y la Media.

- **Moda.** Se utiliza para medir la localización de los datos en una *escala nominal*. Es aquel valor que se repite con más frecuencia en un conjunto de datos, por ejemplo, se está interesado en estudiar los conocimientos y actitudes frente a la sexualidad en un grupo de niñas de un Colegio. Obviamente, es importante para efecto del análisis, la edad de los sujetos en estudio.

Tabla 6.1. Distribución según grupos de edades.

Grupo A. alumnas mujeres

Edad (X)	Frecuencia (f)	f(x)
9	1	9
8	2	16
7	3	21
6	2	12
Total	8	58

La pregunta básica que se hace en el intento de describir globalmente este grupo o colectivo de estudio es ¿cuál es la edad más representativa en el grupo? A simple vista se puede afirmar que la edad representativa es 7 años, ya que la mayoría declaró esa edad; luego *la Moda* es igual a 7 ya que es el valor de la distribución que aparece con mayor frecuencia.

- **Mediana.** Es una medida de tendencia central que divide a cualquier distribución o colectivo en 2 partes iguales o por la mitad y sirve para medir el punto que está al centro de la distribución. Veamos con una ejemplificación en que consiste esta medida.

- Ejemplificación

Calcular la mediana en la información que representa el promedio de días de estadía de los pacientes en el hospital universitario:

Tabla 6.2. Número de días de hospitalización según paciente.

Paciente	Nº días	Paciente	Nº días	Paciente	Nº días
1	1	10	4	19	10
2	2	11	4	20	10
3	2	12	4	21	15
4	3	13	4	22	25
5	3	14	5	23	30
6	3	15	5	24	100
7	3	16	5	25	150
8	3	17	7		
9	3	18	8		

- Análisis de los datos

La mediana es igual a 4: valor que ocupa el lugar 13 en el listado. Este valor se puede interpretar así: la mitad de los pacientes tienen una permanencia de menos de 4 días de hospitalización y la otra mitad más de cuatro días. El número promedio de días de hospitalización de un paciente típico (Media) es de 16.36 días. La Mediana es igual a 4, observe que son dos valores estadísticos muy diferentes. La Media es mucho más sensible a los valores extremos que la Mediana. En este caso, la media está fuertemente influida por dos pacientes que han estado hospitalizados 100 o más días. La mediana, al contrario, está menos determinada por estos.

- Nota metodológica

Usar la Mediana conjuntamente con la Media es un buen recurso metodológico para analizar el comportamiento de un conjunto de datos, por ejemplo, las estadísticas económicas generalmente presentan la mediana de ingresos de un grupo de personas antes que *la media de los ingresos*, ya que el nivel de ingresos puede aparecer artificialmente alto o bajo debido a la presencia de un pequeño segmento de estas personas que pueden ser ubicadas o clasificadas en los extremos, por sus bajísimos o altísimos ingresos.

Para calcular la mediana se tienen que ordenar los datos y determinar el que está a la mitad. Sirve para datos propios de los niveles de medición ordinal, por intervalos y de razón o proporcionales. No es lógico usar ésta medida con variables nominales, porque en este nivel no hay jerarquías, no existe la noción de un escalafón superior o inferior.

La Media. Llamada también media aritmética es la medida de localización o tendencia más utilizada. Con ella se calcula el promedio de los datos. En el lenguaje cotidiano, se le conoce como promedio aritmético. Por ejemplo, de acuerdo a las edades declaradas en la Tabla 6.1 podemos decir que el promedio es de 7.25 años.

La media aritmética se calcula sumando los puntajes de una distribución de valores y dividiéndolos por el número de puntajes u observaciones. Este concepto se resume en la siguiente formula:

$$\text{Cálculo de la Media (X)} = \frac{\sum f X}{n} = \frac{n_1 + n_2 + \dots + n_8}{\sum} = \frac{58}{8} = 7.25$$



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

ceo@catios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

3.1.1.2. Estadísticas para analizar la dispersión o variabilidad de un conjunto de datos

En el análisis de unas características, se necesita algo más que la distribución de los valores de unas respuestas preestablecidas (frecuencias) o el cálculo de unos promedios; también es necesario determinar el grado de variabilidad - diseminación – dispersión - de los datos con respecto a ese promedio. Para tales efectos las medidas de dispersión, tales como el rango, la varianza y la desviación estándar son de extrema utilidad. Estas medidas de variabilidad indican la dispersión y distancias de los valores en la escala de medición.

Rango: Se define como *la diferencia o distancia* entre los valores del dato mayor y el dato menor. Este estadístico mide la dispersión de los valores en una serie. Cuando más amplio sea el rango, mayor será la dispersión de los datos en una distribución.

- Ejemplificación

Supongamos que estamos analizando el número de consultas diarias atendidas por el personal médico de los consultorios del Servicio Médico de la Universidad.

Tabla 6.3. De consultas diarias en los 2 consultorios.

Consultorio A		Consultorio B	
Médico	Nº. Consultas	Médico	Nº. Consultas
A	10	F	28
B	20	G	29
C	30	H	30
D	40	I	31



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

E 50
 X = 30

J

32
X = 30

- **Análisis de los datos**

En ambos consultorios el número promedio de consultas es de 30 pacientes por día. Pero como se logra apreciar el número de pacientes atendidos por cada médico es diferente.

La distribución de las consultas no es idéntica en los dos consultorios: en el Consultorio A, existe una distribución desigual de las consultas, los médicos D y E atienden de dos a cinco veces más pacientes de los que atienden sus colegas A y B.

El rango de consultas en el Consultorio A es de 40 (10 el mínimo y 50 el máximo). En el Consultorio B el "rango" de consultas es de sólo de 4 (32 consultas menos 28 consultas).

- **Desviación estándar o típica.** Estadísticamente, cuando se mide la dispersión lo que se está midiendo es el alejamiento de los valores en relación con la Media. La desviación estándar es la medida de dispersión más usada y se interpreta con relación a la media. Tiene la desventaja que los valores extremos en el conjunto de datos distorsionan su valor.

- **Manejo de estadísticos con el programa SPSS**

En los módulos anteriores se han presentado algunas técnicas de manejo estadístico ejemplificaciones simplificadas y muestras de tamaño menor. Pero en la vida real, como analistas, nos enfrentamos a problemas mucho más complejos que requieren de grandes poblaciones muestrales y del análisis de múltiples propiedades o variables. Los programas de computación especializados facilitan el manejo y análisis de la información.

Todas las técnicas de análisis estadístico que se mencionan en éste módulo se encuentran en el programa SPSS. Por tal razón aquí sólo se hacen referencias a lo que Briones señala como *"la lógica y la oportunidad de su uso en relación con los objetivos de*

investigación y con la naturaleza de la información recogida, y no, en el desarrollo de fórmulas estadísticas...” (Briones, 1988: 21).

A medida que se van desarrollando las técnicas estadísticas, se muestra el procedimiento correspondiente, en una sección dentro del Módulo, que se denomina **Procedimiento SPSS**.

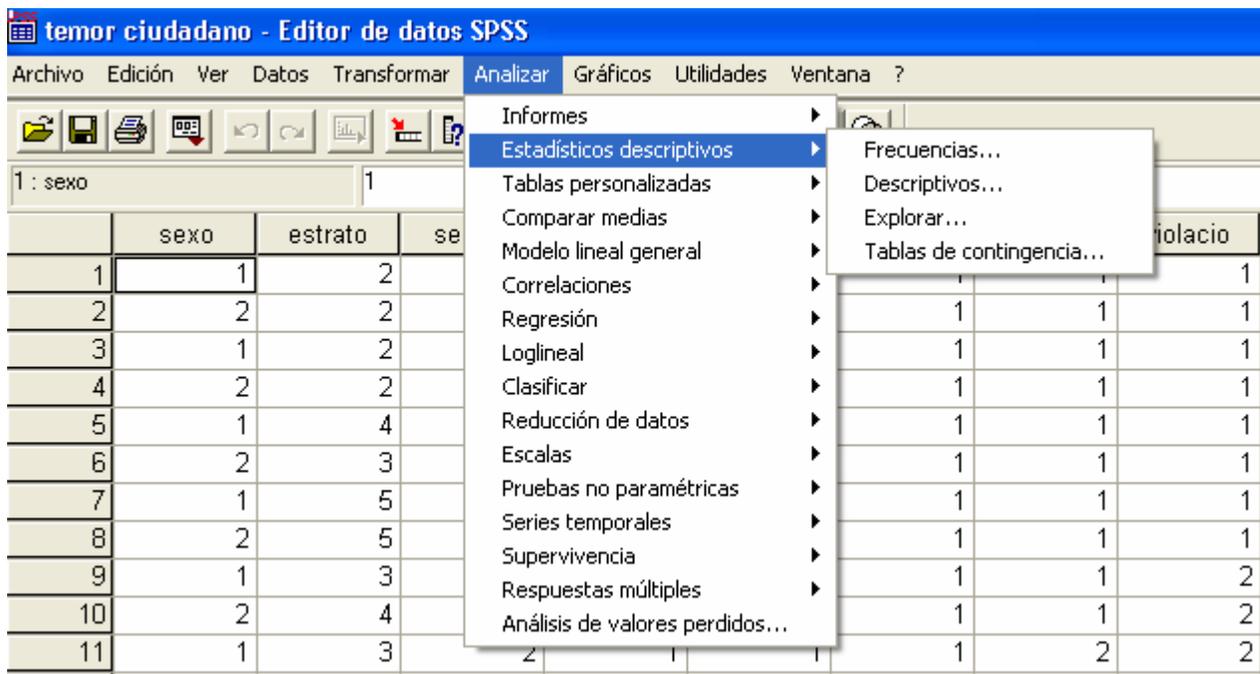
3.1.1.3. Procedimientos generales para la selección de las técnicas estadísticas en el análisis de datos.

Las técnicas de análisis estadístico están agrupadas en el SPSS, en la *Barra Del Menús Principal*, bajo el menú **<Analizar>**.



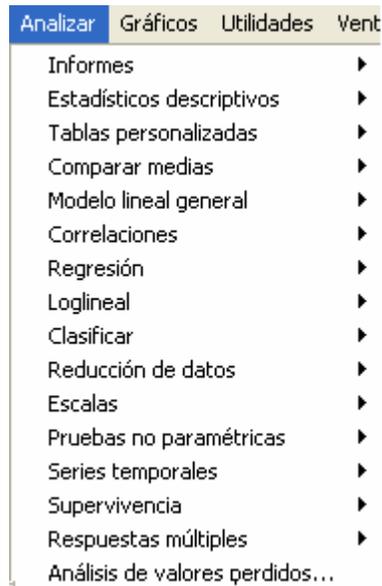
ANALIZAR

Al dar clic sobre este menú aparecerá el listado de técnicas disponibles:

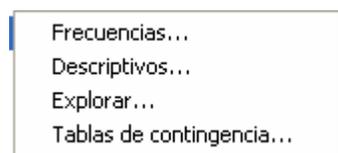


Las opciones del menú Analizar están disponibles sólo si hay un archivo de datos activo o abierto. Estas opciones del menú Analizar ejecutan los procesamientos estadísticos y están organizadas dentro de submenú de acuerdo al tipo de análisis a ejecutar.

El Menú Analizar presenta las siguientes opciones de operaciones de la estadística descriptiva, relacional e inferencia³⁹.



En esta primera fase, nos detendremos en la opción **Estadísticos descriptivos** dando un clic sobre la opción. A continuación, se desplegará una ventanilla que incluye:



La opción Frecuencias es la más utilizada del programa SPSS, dado que no sólo sirve para el cálculo de estadísticos, sino también para depurar datos. Básicamente elabora

³⁹ Las opciones o comandos del menú Analizar están disponibles sólo si hay un archivo de datos activo -abierto.

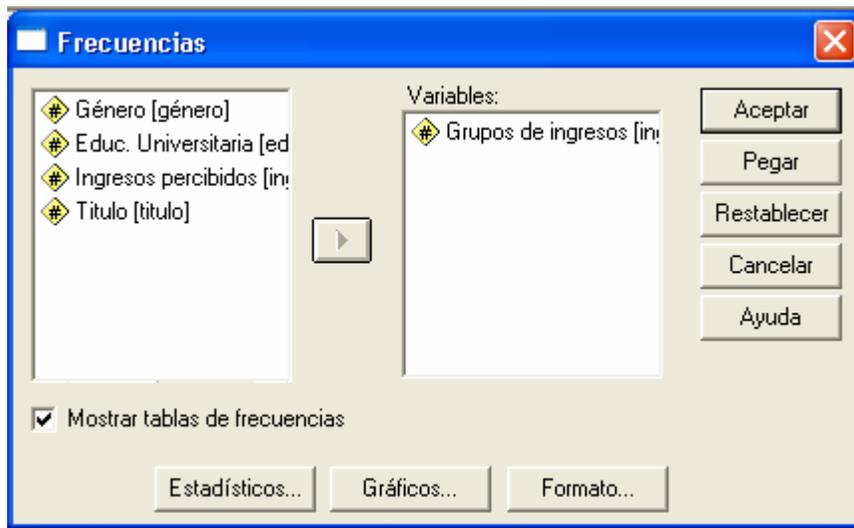
una tabla de frecuencias en donde indica cada categoría que aparece como respuesta, el número de veces que aparece y el porcentaje que representan esas apariciones son respecto al total de casos. Además, con esta opción de estadísticos de frecuencias, se examinan las variables mediante cuatro aspectos descriptivos.

Para obtener estadísticos descriptivos y de resumen opcionales para las variables numéricas utilice el siguiente procedimiento SPSS:

- En la barra principal de Menús del SPSS, seleccione el menú **<Archivos>**
- Active la opción **<Abrir>** elija el submenú **<Datos >**
- Seleccione el archivo **profesional-2003.sav** en el CD-ROM
- En la barra de Menús, elegir el menú **<Analizar⁴⁰>** »»» **<Estadísticos descriptivos>**
- *Seleccionar <Tablas de contingencia>* »»» *Frecuencias...*

Lo anterior abrirá el cuadro de diálogo: Frecuencias, como se muestra en la siguiente figura:

⁴⁰ Cada elemento de la barra de Menú, tiene un menú desplegable. Para acceder al menú respectivo use el puntero del ratón u oprima simultáneamente la tecla ALT más la letra subyacente de la opción.



Si usted quiere calcular algunos estadísticos, *de un clic* en el botón **Estadísticos** y aparecerá a continuación, el cuadro de diálogos Frecuencias: Estadísticos: proceda a seleccionar los estadísticos, por ejemplo, Cuartiles, Desviación típica, Mínimo, Máximo, Media



Frecuencias: Estadísticos

Valores percentiles

Cuartiles

Puntos de corte para 10 grupos iguales

Percentiles:

Añadir

Cambiar

Borrar

Tendencia central

Media

Mediana

Moda

Suma

Los valores son puntos medios de grupos

Continuar

Cancelar

Ayuda

Dispersión

Desv. típica

Varianza

Amplitud

Mínimo

Máximo

E.T. media

Distribución

Asimetría

Curtosis

Al dar un clic en el botón **continuar**, se retorna al cuadro de diálogo Frecuencias. Allí nuevamente se da un clic en el botón **Aceptar**.

A continuación aparecerá el siguiente visor de resultados:

Estadísticos		
Ingresos percibidos		
N	Válidos	1100
	Perdidos	0
Media		26064,20
Desv. típ.		6967,98
Mínimo		7200
Máximo		65500
Percentiles	25	21000,00
	50	26000,00
	75	30375,00

- Análisis de los datos

A partir de la información obtenida de 1100 encuestados, se puede concluir que el promedio de ingresos anuales declarados es de US\$ 26.064; igualmente se observa que existe una gran dispersión en la distribución de los ingresos – US\$ 7.200 como mínimo y US\$ 65.500 como máximo. Se puede afirmar, a partir de los datos que el 25 por ciento de los encuestados tiene ingresos iguales o inferiores a US\$ 21.000 y que el 50 por ciento tiene ingresos iguales o superiores a US\$ 26.000.

- Nota metodológica

El cuadro de diálogo **Frecuencias: Estadísticos** presenta cuatro opciones de manejo de estadísticos de localización o distribución:

1. Valores de percentiles. Puede elegir una o más de las siguientes opciones:

Cuartiles. Muestra los percentiles 25, 70 y 75.

Puntos de corte para n grupos iguales muestra los valores del percentil que divide la muestra en grupos de casos de igual tamaño. El número de grupos predeterminado es de 10. De modo opcional, puede solicitar un número de grupos diferente. Introduzca un entero positivo entre dos y cien por ejemplo, si introduce 4, se mostrarán los cuartiles.

Percentil: Valores en los percentiles especificados por el usuario. Introduzca un valor de percentil entre cero (0) y cien (100) y dé un clic en el botón *Añadir*.

2. Dispersión: Puede elegir una o más de las siguientes opciones:

Desviación estándar o típica: Estimación de la cuantía en que se diferencian las observaciones respecto a la media, expresada en las mismas unidades que los datos.

Varianza: Estimación de la cuantía en que se diferencian las observaciones respecto a la media, es igual al cuadrado de la desviación típica.

Amplitud: Diferencia entre los valores mayor (máximo) y el menor (mínimo).

Mínimo: El valor más pequeño.

Máximo: El valor más grande.

E.T. Media: Error típico de la media. Una estimación de la variabilidad muestral de la media.

3. Tendencia central: Puede elegir una o más de las siguientes opciones:

Media: La media aritmética o promedio.

Mediana: La mediana está definida por el valor por debajo del cual se encuentran la mitad los casos. Si hay un número de casos par, la mediana es la media de los dos casos centrales cuando están ordenados en orden ascendente.

Moda: el valor que ocurre más frecuentemente. Si diversos valores están empatados en la frecuencia más alta sólo aparecerá el valor más pequeño de la variable.

Suma: la suma de todos los valores.

4. Distribución: puede elegir una o más de las siguientes opciones:

Asimetría: Índice del grado en que la distribución no es simétrica. También se muestra el error típico del estadístico de asimetría.

Curtosis: estimación del grado en que las observaciones se agrupan alrededor del punto central, también se muestra el error típico del estadístico de Curtosis.

Para datos agrupados, también se encuentra disponible la siguiente opción:

3.1.2. Técnicas de representaciones gráficas de las distribuciones de frecuencias.

Los gráficos o figuras constituyen uno de los medios más usados y más difundidos para la presentación y análisis de la información cuantitativa. Esto se debe al hecho de que las ideas presentadas gráficamente son entendidas con mayor rapidez que las explicaciones textuales o numéricas.

3.1.2.1. Tipos de gráficos

La mayoría de las personas han visto o analizado y hasta construido figuras o gráficos y posiblemente han notado que hay una gran variedad de ellos. Una presentación detallada de todos los tipos excede los propósitos de este texto, por ello sólo se hará referencia a los de mayor uso.

- Histogramas de frecuencia: Son gráficos que presentan la información de cada una de las categorías de la variable la forma de rectángulos proporcionales. Se utilizan generalmente para representar datos de una variable discreta.
- Polígonos de frecuencias: Son gráficos en la forma de una serie de líneas rectas conectadas entre sí y que unen unos puntos medios ó de intervalo a lo largo del eje horizontal. Se utilizan para representar distribuciones de frecuencias propias de un nivel de medición por intervalos.
- Gráficos de barras: Constituyen uno de los tipos más simples y quizás de mayor uso. Resultan de especial utilidad en la presentación de datos cualitativos o cuantitativos. Se emplean tipos muy variados de gráficos de barras, sin embargo todos reúnen ciertas características o requisitos esenciales.
- Gráficos de círculos o de pastel: En este tipo el área de cada una de los sectores refleja la importancia de la categoría que representa. Se trata de un gráfico en dos dimensiones. El gráfico de pastel consiste en un círculo el cual se divide en tantos



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

sectores como categorías se tienen de manera que el área que corresponde a cada categoría sea proporcional a su importancia relativa o porcentual

3.1.2.2. Limitaciones en el uso de los gráficos o figuras

La utilización de figuras o gráficos tiene ciertas limitaciones que es conveniente tener en cuenta:

- No pueden incluir tanta información como las tablas. Sólo se puede presentar a la vez una cantidad limitada de datos ya que demasiada información y/o variedad atenta contra su utilidad, al dificultar su lectura.
- En las figuras o gráficos se representan valores aproximados mientras que las tablas permiten incluir la información con toda exactitud.
- La presentación tabular y gráfica es complementaria al texto pero no lo reemplaza. Por ejemplo, para ilustrar el texto, igualmente para llamar la atención y despertar el interés por los datos que son presentados así como para reforzar las argumentaciones y conclusiones a las cuales se ha llegado.

- Procedimiento SPSS

En la barra de Menús del SPSS, seleccione el menú Gráficos⁴¹ » » » Barras » » » Gráficos de barras » » » seleccione el tipo de grafico⁴²:

También se pueden obtener gráficas con algunas opciones del Menú Analizar » » » Estadísticos descriptivos » » » Frecuencias » » » Botón Gráficos

⁴¹ Este menú ofrece opciones de figuras tales como: barras, líneas, áreas, sectores, histograma con sus respectivos cuadros de dialogo: Grafico de...

⁴² Por ejemplo, el SPSS puede diseñar nueve tipos de gráficos de barra.

3.1.3. La interrelación entre variables

En ocasiones, un investigador requiere que su análisis vaya más allá de la mera descripción de una variable; la misma complejidad de la realidad sociedad exige estudiar, por ejemplo, cómo un proceso esta asociado a otros o cómo varía en función de otros. Cada día, los analistas e interventores sociales toman decisiones que están basadas sobre predicciones de eventos futuros. Para hacer estos pronósticos ellos se basan en la relación -intuitiva y calculada- entre lo que se conoce y lo que se va a estimar. Si los administradores pueden determinar cómo está relacionado lo conocido con el evento futuro, pueden ayudar considerablemente al proceso de toma de decisiones. Este es el tema de este Capítulo: *cómo determinar la relación entre variables*.

En el módulo anterior, se expusieron diversos procedimientos que permiten la descripción de colectivos. En el presente Módulo y en el siguiente, usted, encontrará diversas técnicas que permiten analizar determinaciones y relaciones entre variables.

- Ejemplificación

Supongamos que se estudian las diferencias de opinión que tienen personas de distinto sexo acerca de cuáles serían las soluciones al problema de la inseguridad ciudadana.

Examinemos la relación entre las variables *género y posibles soluciones a la violencia*. A partir de ésta se puede formular una hipótesis: con mayor frecuencia que los hombres, las mujeres proponen que un aumento en la dotación policial, es una solución a la violencia que vive la Ciudad.

De hecho, el enunciado anterior señala una relación. El problema metodológico surgirá entonces en el momento en que queramos confirmar nuestra afirmación: *¿Cómo*

comprobar la relación entre las dos variables género y soluciones posibles a la inseguridad y violencia? Es decir, cómo una explica, determina u afecta a la otra. O, cómo una varía en función de la otra. Una posible solución metodológica es la de construir una tabla de contingencia o de tabulación cruzada.

3.1.3.1. Tablas de contingencia y medidas de asociación.

Una de las técnicas más frecuentemente utilizadas en la investigación social es la clasificación de datos en tablas según uno o más criterios (variables). Veamos con un ejemplo de una tabulación cruzada como se clasifican los datos de dos variables: *género y preferencia por un tipo de solución al problema de la inseguridad ciudadana.*

Tabla 6.4. Alternativas desde la perspectiva de género para disminuir la delincuencia en la ciudad de Medellín.

	Hombres	Mujeres	Total
Aumentar dotación policial	38%	17%	26%
Participación comunitaria	32%	29%	31%
Otras soluciones	19%	46%	34%
N.S./ N. R.	11%	8%	9%
Total	100%	100%	100%

3.1.4.2. Análisis de datos tabulados

La lectura de los datos de una tabla depende de la hipótesis de trabajo, lo cual determina el tipo de cálculo porcentual que se requiere. Los porcentajes en la Tabla son calculados teniendo como base el total de la frecuencia en cada categoría de la variable independiente (género), así la suma de los porcentajes deberá sumar 100 por ciento en la base de la variable independiente. Sin embargo, uno se podrá preguntar, *¿porque no hacer el cálculo en la dirección de las categorías de la variable dependiente -alternativas de soluciones a la delincuencia-?*

La respuesta es muy sencilla y a la vez muy importante ya que la mayoría de los programas estadísticos de computador generan tablas de contingencia con porcentajes calculados en ambas direcciones –**horizontales - verticales** y es el Analista quién debe decidir cuál es la dirección correcta en la lectura de la Tabla.

La tabulación cruzada es una técnica estadística también denominada *Tabla de Contingencia*, en la que se presentan los valores de las frecuencias conjuntas de dos o más variables. Cuando se forman con dos variables reciben el nombre de *tablas de dos entradas*; cuando son más de dos las variables involucradas, son tablas de *n entradas*, donde n es el número de variables involucradas. A cada tabla obtenida del cruce de dos o más variables se le conoce como *subtabla*, dado que en realidad es parte de la tabla principal formada por las dos primeras variables. En ella se calcula la distribución de frecuencias de una variable para cada una de las categorías o clases en las que se divide la otra variable con la cual se cruza. Los porcentajes en una tabla cruzada, que expresan frecuencias relativas, pueden calcularse en dirección vertical u horizontal. *Los porcentajes verticales* se calculan sobre el total de elementos de la muestra que pertenecen a la categoría indicada en cada columna. *Los porcentajes horizontales* se calculan sobre el total de cada fila. La frecuencia de cada fila o columna se denomina *frecuencia marginal*.

Se supone que la variable independiente es el factor causal o determinante que tiene el mayor peso en las posibles variaciones de la variable dependiente. La decisión acerca de cuál es *la variable independiente o explicativa* y cuál es la dependiente o explicada parece fácil, pero en ciertas ocasiones tiende a ser de tal complejidad que genera traumatismos, especialmente cuando se está trabajando con programas como el SPSS.

Por ejemplo, en la relación entre *género y alternativas de solución*⁴³, suena sin sentido considerar como variable independiente a *las alternativas de solución* y parecería sin sentido pensar que esta variable pueda influir al *género*. Sin lugar a dudas aquí la variable independiente es el *género*, del cual nosotros hemos hipotetizado que afecta o determina la variable dependiente (*alternativas de solución*)

Si el sentido común, a veces limitado, no nos ofrece una pista acerca de *cuál es cuál*, un criterio metodológico adoptable es el de establecer la secuencia temporal entre las dos variables:

¿Cuál de las dos variables precede en el tiempo a la otra?

Procedimiento SPSS

En el SPSS el submenú *<Tablas de contingencia>* permite la tabulación cruzada y calcula sus estadísticas. En esta operación, se puede utilizar cualquier tipo de variables: *cuantitativas discretas, cuantitativas continuas, cualitativas nominales, cualitativas ordinales*.

Todas las estadísticas que calcula *<Tablas de contingencia>* son *medidas de asociación* o, en su defecto, **de independencia**; pero la negación de una lleva la afirmación de la

⁴³ Estudio sobre la inseguridad ciudadana.



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

otra; si dos variables no están asociadas, son independientes, o si dos variables no son independientes están asociadas.

- Ejemplificación

Sistematizar y analizar la información del estudio sobre *la apreciación de los colombianos sobre la crisis política que vive el país*.

- En la barra principal de Menús del SPSS, seleccione el menú **<Archivos>**
- Active la opción **<Abrir>**»»»»»»»» elija el submenú **<Datos >**
- Seleccione el archivo **situación-2000.sav** en el CD-ROM
- En la barra de Menús, elegir el menú **<Analizar>**»»»»»»»» **<Estadísticos descriptivos>**
- *Seleccionar <Tablas de contingencia>*
- Seleccione una variable independiente de la base de datos **Situación-2000.sav**, por ejemplo –*sexo*- y “*crúcela*” con la variables “sensación de inseguridad”
- Proceda a cruzar las variables con el siguiente procedimiento:
- Transfiera la variable independiente *sexo* a “Filas”:
- Transfiera la variable dependiente *sensación de inseguridad* a “Columnas”
- En el botón Cásillas defina si el cálculo de los datos se va a hacer por **filas, columnas o sobre el total de casos**



A continuación, en el visor de resultados aparecerá la siguiente tabla:

Tabla de contingencia sexo * sensacion de inseguridad

			sensacion de inseguridad		Total
			sentirse inseguro	sentirse seguro	
sexo	femenino	Recuento	11	1	12
		% de sexo	91,7%	8,3%	100,0%
	masculino	Recuento	4	9	13
		% de sexo	30,8%	69,2%	100,0%
Total		Recuento	15	10	25
		% de sexo	60,0%	40,0%	100,0%

- Análisis de los datos

Del total de personas entrevistadas de sexo femenino, una gran mayoría manifiesta sentirse insegura (91.7%); las personas de sexo masculino, en cambio, manifiestan en menor medida dicho temor: Sólo un 30.8% de los entrevistados se siente inseguro.

- Nota metodológica

Con el procedimiento <Tablas de contingencia> se puede utilizar cualquier tipo de variables: Cuantitativas o cualitativas, sean ordinales o nominales. Por supuesto, las estadísticas que se calculan tienen validez sólo para las variables adecuadas, por ejemplo, el coeficiente de correlación entre sexo y edad no tiene ningún sentido, ya que el sexo es una variable cualitativa nominal y la edad es cuantitativa continúa.

- Control del efecto de las variables intervinientes en una relación original.

Uno de los propósitos de los analistas sociales que utilizan los diseños cuantitativos es el de tratar de aislar o circunscribir, en lo posible, aquellos elementos, hechos o factores sociales que median un proceso social. En términos de variables, lo que se trata es de intervenir o aislar aquellas "*variables intervinientes*"⁴⁴ (Escalante, Carlos, 1987: 83) que pueden ser consideradas causas de los cambios de la variable dependiente.

Supongamos que queremos determinar si un programa de alcoholismo "cura" alcohólicos. Para esto debemos estar seguros que "la curación" es debida al Programa y no a otros factores intervinientes o asociados.

En los diseños cuantitativos, se dice que existe una conexión causal entre dos variables cuando:

- Las dos varían o covarían, esto es, que si una varía la otra también debe hacerlo;
- Que una precede en el tiempo a la otra, es decir, que existe una secuencia temporal entre ellas;
- La relación entre las dos no es el resultado de otra variable que esta relacionada a ellas;
- Existe un sólido argumento teórico que permite asegurar que una de las variables es la determinante de la otra.

Gracias al uso de controles estadísticos es posible acercarnos con cierto grado de confiabilidad a cuál es realmente la variable más determinante o causal del proceso que se está analizando y, de paso, eliminar la posibilidad que la relación encontrada entre la variable independiente y dependiente sea el resultado de otra variable asociada a ellas.

⁴⁴ El concepto de variable interviniente o de control presupone que la relación entre dos variables esta mediada por otra (u otras) que transportan los posibles efectos de la primera.

Si la relación se mantiene *inmodificada o sostenida* dentro de los subconjuntos, pese a la existencia de la variable interviniente, se puede llegar a afirmar que se ha eliminado una posible causa alternativa en *la relación original* observada. Pero sí la determinación entre las variables independiente y dependiente se mediatiza o desaparece, nosotros sabremos que la relación entre las variables originales no es tan determinante y lo que realmente explica la situación es la relación secundaria.

Tabla A: Soluciones ante el aumento de la violencia según género

	Hombres	Mujeres	Total
Aumentar dotación			
policial	38%	17%	26%
Participación Comunitaria	32%	29%	31%
Otras soluciones	19%	46%	34%
N.S./ N. R.	11%	8%	9%
	100%	100%	100%

En la tabla *Soluciones ante el aumento de la violencia según género*, se aprecia que un 38 por ciento de las personas clasificadas como hombres se inclina por la solución del aumento de la dotación policial, en cambio sólo un 17 por ciento de las mujeres mostró igual preferencia.

Si se analizan las opiniones de los encuestados, al introducir una nueva variable interviniente: *nivel de escolaridad* encontramos que las apreciaciones por el tipo de solución de aumentar de la dotación policial varían sustancialmente.

Veamos que nos muestra la siguiente tabla

Tabla B: Soluciones ante el aumento de la violencia según género

	Alto nivel de escolaridad		Bajo nivel de escolaridad	
	Hombres	Mujeres	Hombres	Mujeres
Aumentar dotación policial	36%	31%	27%	9%
Participación comunitaria	32%	29%	30%	40%
Otras soluciones	20%	32%	33%	36%
N.S./ N. R.	12%	8%	10%	15%
	100%	100%	100%	100%

La lectura comparada de las tablas A y B muestra el efecto de la variable interviniente - *nivel de escolaridad*- Limitemos ahora nuestro análisis a la primera categoría de la variable *soluciones a la violencia ciudadana*: Una primera apreciación señalaría que el nivel de escolaridad muestra un mayor peso, que la variable *género*, en la conformación de un tipo de opinión sobre *posibles soluciones a la violencia ciudadana*.

Una segunda mirada, más bien cualitativa, muestra que las tendencias originales tienden a mantenerse iguales si bien la introducción de la variable interviniente hace variar los guarismos en las dos categorías de la variable nivel de escolaridad

Específicamente aquellas personas de sexo masculino que declaran un mayor nivel de escolaridad, siguen mostrando una mayor inclinación que las mujeres por la solución del aumento de la fuerza policial en la ciudad, al igual en las personas clasificadas con un bajo nivel de escolaridad.

En síntesis, la introducción *una variable interviniente* puede provocar que *la relación original se sostenga, disminuya o aumente*. En el primer caso, si la relación original se sostiene se puede concluir que la variable interviniente no tiene ningún efecto. En las otras dos situaciones, se puede afirmar que *la variable interviniente* afecta la relación

original, y la causa de esto puede ser que ésta antecede tanto como factor causal y temporal a ambas y/o, está relacionada tanto con la variable independiente como con la variable dependiente.

El proceso de agregar nuevas variables intervinientes puede llevar a producir una serie de efectos o situaciones en la relación original de las variables consideradas principales:

- Replicación: la relación original permanece inalterada, así se repitan en diferentes situaciones espacio-temporales;
- Explicación: la relación original es eliminada en los subconjuntos de datos -relación espuria-
- Interpretación: ocurre cuando la relación entre los subconjuntos es reducida o eliminada sin embargo es posible pensar que la variable de control interviene entre la variable independiente y la variable dependiente;
- Especificación: es el término usado para referirse a la situación donde la relación entre los subconjuntos de variables difiere uno de otro; es decir, la relación original puede aumentar sustantivamente en un subconjunto, por ejemplo, entre los hombres y desaparecer entre las mujeres.

- Procedimiento SPSS

- En la barra principal de Menús del SPSS, seleccione el menú **<Archivos>**
- Active la opción **<Abrir>**»»»»»» elija el submenú **<Datos >**
- Seleccionar archivo **situación-2000.sav** en el CD-ROM
- En la barra de Menús, elegir el menú **<Analizar>**»»»»»» **<Estadísticos descriptivos>**
- *Seleccionar <Tablas de contingencia>*
- En el cuadro de diálogo Tablas de contingencia »»»»»» seleccionar la variable *sexo* y transferirla a la ventanilla *Filas* y la variable *diálogos en medio de la guerra* a la ventanilla *Columnas*

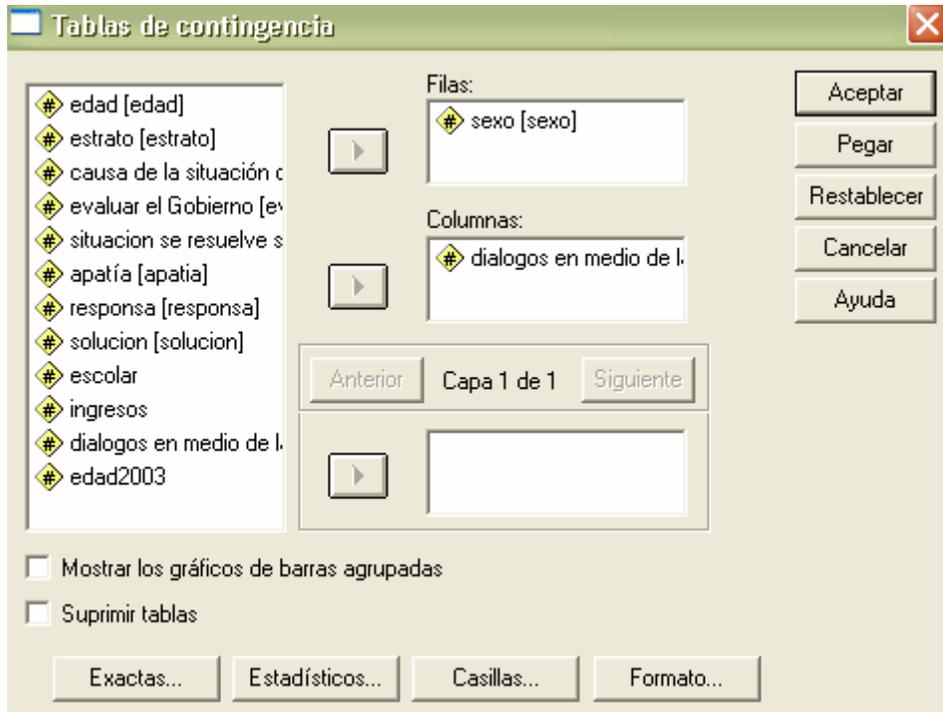


UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

ceo@catios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775



En el visor de resultados, se pueden observar lo siguientes datos:

Tabla de contingencia sexo * dialogos en medio de la guerra

			dialogos en medio de la guerra					Total
			muy en desacuerdo	desacuerdo	no sabe	acuerdo	muy de acuerdo	
sexo	femenino	Recuento	4	4	2	1	1	12
		% de sexo	33,3%	33,3%	16,7%	8,3%	8,3%	100,0%
	masculino	Recuento	1	3	3	2	4	13
		% de sexo	7,7%	23,1%	23,1%	15,4%	30,8%	100,0%
Total		Recuento	5	7	5	3	5	25
		% de sexo	20,0%	28,0%	20,0%	12,0%	20,0%	100,0%

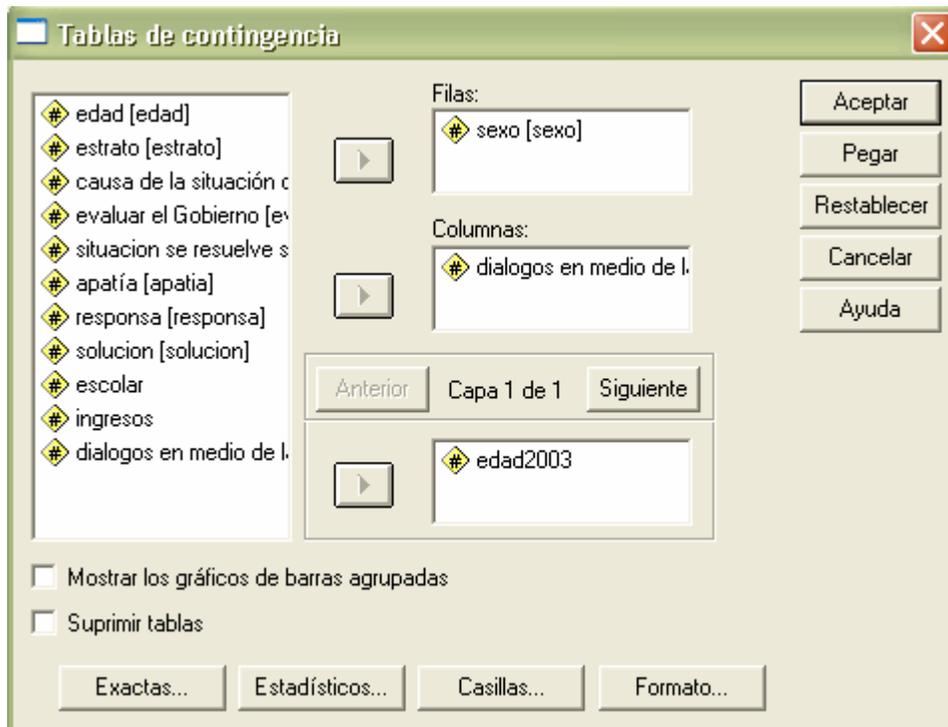
- Análisis de los datos

Con respecto a la opinión de los hombres, se puede apreciar que muestran una mayor predisposición al diálogo en medio de la guerra, un 46.2% de los entrevistados están de acuerdo; en cambio, el 66,6 por ciento de las mujeres entrevistadas muestran un mayor desacuerdo con esta posibilidad.

A la relación original planteada se le introduce una tercera variable interviniente *edad2003* para responder a la inquietud si es el género o la edad la que realmente explica la situación. Si la relación se mantiene *inmodificada o sostenida* dentro de los subconjuntos, pese a la existencia de esa variable interviniente, se puede llegar a afirmar que se ha eliminado una posible causa alternativa en *la relación original* observada. Pero, sí la determinación entre las variables independiente y dependiente desaparece, nosotros sabremos que las dos variables no están relacionadas y la que realmente explica la situación es la nueva relación.

En el cuadro de diálogo Tablas de frecuencia, seleccionar la variable recodificada Edad 2003⁴⁵ y transferirla a la 3ª ventanilla –debajo de “Capa 1 de 1”. A continuación dar un clic en el botón **Aceptar**.

⁴⁵ Categorías y códigos de la variable edad2003: jóvenes = 1 y adultos = 2



A continuación, en el visor de resultados aparecerá la siguiente tabla:

Tabla de contingencia sexo * dialogos en medio de la guerra * EDAD2003

EDAD2003				dialogos en medio de la guerra					Total
				muy en desacuerdo	desacuerdo	no sabe	acuerdo	muy de acuerdo	
1	sexo	femenino	Recuento	3	3		1		7
			% de sexo	42.9%	42.9%		14.3%		100.0%
	masculino	Recuento		3	2	1	2	8	
		% de sexo		37.5%	25.0%	12.5%	25.0%	100.0%	
	Total	Recuento	3	6	2	2	2	15	
		% de sexo	20.0%	40.0%	13.3%	13.3%	13.3%	100.0%	
2	sexo	femenino	Recuento	1	1	2		1	5
			% de sexo	20.0%	20.0%	40.0%		20.0%	100.0%
	masculino	Recuento	1		1	1	2	5	
		% de sexo	20.0%		20.0%	20.0%	40.0%	100.0%	
	Total	Recuento	2	1	3	1	3	10	
		% de sexo	20.0%	10.0%	30.0%	10.0%	30.0%	100.0%	

- Análisis de los datos

Se puede apreciar que en los jóvenes⁴⁶ (código1), la relación original persiste, tanto los hombres como las mujeres mantienen sus opiniones. En cambio, los adultos (código 2) manifiestan opiniones encontradas: Las mujeres adultas manifiestan un menor desacuerdo y los hombres un menor acuerdo. Se puede concluir, en términos generales, que el género es la variable explicativa.

- Reagrupamiento de los valores de una variable

En el supuesto que la información, tanto para la variable independiente como para la variable dependiente, sea de tipo interval, se puede anticipar un gran número de valores para cada variable. Una tabla de contingencia con muchas categorías o valores sería de difícil manejo, por no decir de imposible lectura. Para evitar esto, generalmente se debe

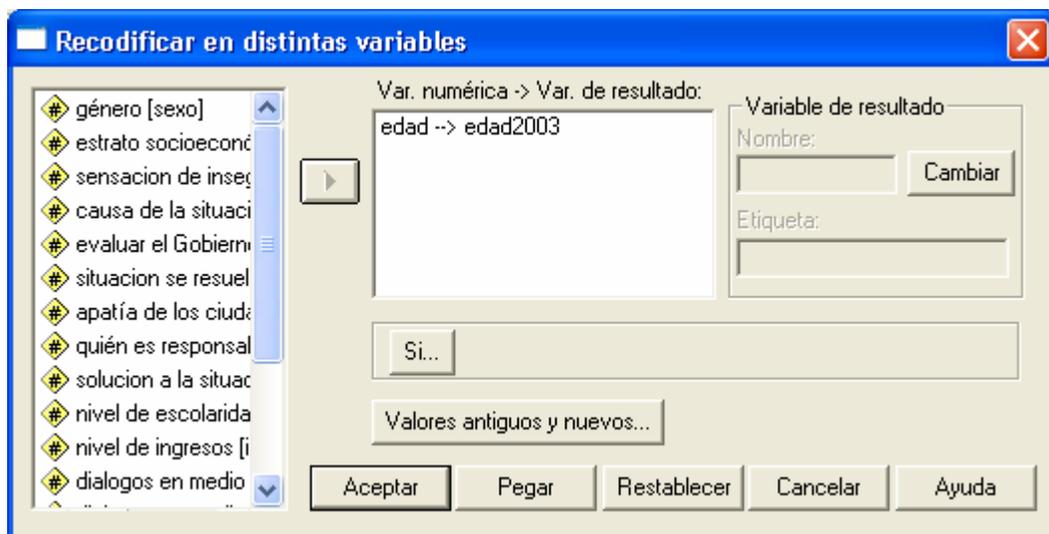
⁴⁶ Código 1= personas menores de 30 y menos años y Código 2 = personas de mas de 30 años.

resumir o reagrupar⁴⁷ la información, lo que se trata es transformar a un número limitado de categorías (generalmente 8 o 10). En síntesis, la recodificación de valores significa cambiar su valor actual por uno asignado por el analista.

⁴⁷ Recodificación de los valores en rangos iguales

- Procedimiento SPSS

- En la barra principal de menús del SPSS, seleccione el menú <Archivos>
- Active la opción <Abrir> »»»»»»» elija el submenú <Datos >
- Seleccione el archivo **situación-2000.sav** en el CD-ROM
- En la barra de Menús, elegir el menú <Transformar> »»»»»»»→ Recodificar
- Elegir la opción <En distintas variables>
- Aparece el cuadro de diálogo Recodificar en distintas variables:
- Se selecciona de la ventanilla con el listado de variables, la variable *edad* y se transfiere a la ventanilla central
- La variable aparecerá así en la ventanilla central; edad →?
- Escriba el nuevo nombre de la variable recodificada: *edad2003* en la casilla Nombre de la sección “Variable de resultado”
- De un clic en el botón C
- En la ventanilla central aparecerá así: edad → edad2003





UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

- Definido el nuevo nombre se da clic en el botón **Valores antiguos y nuevos...**

- A continuación se abrirá el siguiente cuadro de diálogo:

En este cuadro de diálogo se sigue la siguiente secuencia de procedimientos:

- En la sección Rango situarse en la ventanilla izquierda y digitar el rango mínimo: 1
- En la ventanilla derecha, digitar el rango máximo: 2
- En la ventanilla Valor nuevo situarse en la casilla Valor y escribir el nuevo valor: 1
- De un clic en el botón Añadir
- Los nuevos valores recodificados aparecerán en la sección Antigo → Nuevo
- Repita los cuatro (4) pasos anteriores con los valores originales y el nuevo.
- Dé un clic en el botón continuar y luego en el botón Aceptar

ceo@catios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

- En la matriz de datos aparecerá la nueva variable *edad2003*

3.1.4. Inferencia estadística

El objetivo de una investigación no se reduce al trabajo descriptivo solamente, en muchos casos, lo que se pretende es generalizar los resultados obtenidos a partir de una muestra de la población.

"Los datos casi siempre son recolectados de una muestra y sus resultados estadísticos se denominan estadígrafos, por ejemplo, la media o la desviación estándar de la distribución de una muestra son estadígrafos. A las estadísticas de la población o universo se les conoce como parámetros. Los parámetros no son calculados, por que no se recolectan datos de toda la población, pero pueden ser inferidos de los estadígrafos, de ahí el nombre de estadística inferencial. (Hernández, Roberto, Et, Al., 1991:373)

La estadística inferencial sirve fundamentalmente para dos procesos: *Pruebas de hipótesis y estimativos de parámetros poblacionales.*

En este Capítulo, se indican algunas técnicas estadísticas en la que se usan las características de las muestras para probar una hipótesis que se ha formulado de un parámetro de población. Los procedimientos estadísticos a utilizar dependerán, obviamente, del nivel de medición de las variables independiente y dependiente. Cuando las dos variables son nominales u ordinales, generalmente se utilizan las llamadas tablas de contingencia. Si una es nominal u ordinal y la otra de intervalos o proporcional, *los test de diferencias de medias o de análisis de varianza* son las técnicas preferidas. Y cuando ambas variables son de intervalo o proporción, el análisis de regresión o correlación son los más adecuados.

En los procesos de medición de relaciones entre variables y prueba de hipótesis hay cinco preguntas que deben ser tenidas en cuenta: *¿Hay relación entre las variables*

independiente y dependiente? ¿Cuál es la dirección o forma de la relación? ¿Cuán fuerte es la relación? ¿Es la relación estadísticamente significativa? ¿Es la relación de tipo causal?

Supóngase que se están evaluando terapias de rehabilitación en pacientes fármaco dependiente. Antecedentes sobre esos procesos de rehabilitación muestran que alrededor del 42 por ciento de los drogadictos que han asumido individualmente su terapia (tratamientos tradicionales) reinciden en estados de drogadicción.

Se hipotetiza que sólo el 28 por ciento de los drogadictos que participan en un programa innovador de rehabilitación reinciden en el consumo de alucinógenos. ¿Es posible, a partir de estas diferencias, afirmar que la metodología innovadora de dicho programa es la más adecuada para el tratamiento de la drogadicción?

No se puede aceptar o rechazar una hipótesis acerca de un parámetro de una población por simple intuición. En lugar de ello, se necesita aprender a decidir objetivamente, bajo las bases de la información muestral, cuando aceptar o rechazar una intuición.

Se hipotetiza un cierto valor para la media de la población de reincidentes (42%). Para probar su validez, se obtienen datos muestrales y se determina la diferencia entre el valor de la hipótesis y el valor real de la media muestral. Luego se juzga si la diferencia es significativa.

Mientras menor sea la diferencia, mayor será la probabilidad de que el valor hipotético para la media esté correcto. Pero, en no todas las ocasiones, la diferencia entre el parámetro hipotético de la población y el estadístico muestral es tan grande que automáticamente se rechace la hipótesis, ni tan pequeña que se acepte inmediatamente.

Por tanto, en *pruebas de hipótesis* como en la mayoría de las decisiones significativas, las soluciones precisas son la excepción, no la regla.

La ejemplificación anterior ilustra la necesidad de ir más allá de la sola descripción de las posibles diferencias cuantitativas entre dos resultados -terapias de rehabilitación-. Cabe preguntarse si las diferencias encontradas son un reflejo real o son simplemente *unas diferencias aritméticas no significativas*. Son en casos como éstos, donde las llamadas *pruebas de significación estadística* pueden ayudar a responder a este tipo de interrogantes. Estas herramientas estadísticas son útiles para aceptar o refutar las posibles diferencias o variaciones que se puedan presentar en un conjunto de unidades de análisis.⁴⁸

En síntesis, se trata de obtener conclusiones que sean generalizables a una población a partir de una información muestral resumida en un conjunto de estadísticos llamados inferenciales. Frecuentemente, por razones de tiempo y costo u otras consideraciones, no es posible examinar todas las unidades de análisis que componen el universo del estudio por eso se deben tomar las medidas del caso para que los resultados de la muestra que se observa y se analiza reflejen realmente las características de la población total.

Para lograr este objetivo es necesario conocer algunas características de la población en estudio para determinar el tipo de muestreo que requerimos. Luego, de obtenida de debe efectuar una prueba de hipótesis sobre los estadísticos obtenidos con la muestra (media y desviación estándar), con respecto a los parámetros poblacionales, para determinar su consistencia y estabilidad cuando el experimento se repita, con el fin de hacer algunas generalizaciones.

⁴⁸ Individuos, grupos, instituciones, hechos, eventos, divisiones político-administrativas o geográficas, etc.

De otro lado, no olvidemos nuestra preocupación –en el ejemplo anterior- acerca de sí la tasa de reincidencia en el programa innovador es realmente concluyente y por lo tanto diferente a las tasas del *programa tradicional*. Para responder el interrogante planteado el primer paso a desarrollar será el de definir *la población muestral de ambos grupos de rehabilitación*. Lamentablemente no hay técnica que nos defina una muestra poblacional cien por cien confiable; sin embargo, *una prueba de significancia estadística* puede permitirnos estimar la probabilidad de que el tamaño de las diferencias encontradas entre dos grupos se deba al azar. Con el fin de entender el concepto de significancia estadística, se hace necesario comprender la lógica del muestreo.

3.1.4.1. Prueba de hipótesis

Una vez recogida la muestra probabilística, la importancia de ella radica en que permita predecir el comportamiento de la población. Recordemos que un buen modelo cuantitativo es aquel que logra eliminar la necesidad de explicaciones alternativas para el efecto causado, si es que lo hay, en la variable dependiente, ya que las muestras variarán desde su población por azar. Como en la mayoría de los casos no se conoce la media poblacional verdadera, podemos tener un valor aproximado como resultado del conocimiento o de la experiencia acerca del objeto de estudio.

El procedimiento para la contrastación de hipótesis comprende las siguientes etapas:

- Planteamiento de una hipótesis nula (H_0) o hipótesis a contrastar.
- Formulación de una hipótesis alternativa de investigación (H_i).
- Elección del riesgo o nivel de significancia que prácticamente es el complemento de lo que se llama nivel de confianza. Normalmente $\alpha = 0.10$ ó 0.05 .
- Contraste de hipótesis. Comparar el valor encontrado del estadístico de prueba obtenido con el valor crítico y se toma *la decisión de aceptar o rechazar la hipótesis nula*.

La prueba de hipótesis comienza con una afirmación o propuesta tentativa, que se hace acerca de un parámetro de una población, llamada hipótesis.⁴⁹ Luego se recogen datos

⁴⁹ En la estadística inferencial hay dos tipos de hipótesis: Una. Es la hipótesis de investigación que señala la posible existencia o relación entre dos hechos. La otra, es la hipótesis nula, que se define para que el analista evalúe su hipótesis de investigación.



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

muestrales, se producen estadísticos muestrales y se usa esta información para decidir qué tan probable es, que la hipótesis del parámetro de población esté correcta.

3.1.4.2. Selección del test estadístico

Tengamos en cuenta que la hipótesis estadística es una conjetura o suposición concerniente a las características de la población; antes de aceptarla o rechazarla el investigador debe probar su validez. Ante la tarea de probar una hipótesis, se usa una muestra de acuerdo a la teoría de las probabilidades. El resultado de la prueba conducirá a unos estadísticos, tales como *t de Student* o el *Chi-Cuadrado*, con los cuales el investigador podrá decidir sobre la validez de su propuesta.

- Valores de los niveles de significancia

El objetivo de una prueba de hipótesis es hacer un juicio sobre la **diferencia** entre la estadística muestral y el parámetro hipotético de la población. Como se ha adelantado, el siguiente paso después de definir la hipótesis de investigación y la hipótesis nula es decidir qué criterio –valor- usar para aceptar o rechazar la hipótesis nula.

Supongamos que obtenemos un estadístico muestral (una media aritmética), entonces debemos decidir con qué porcentaje de confianza este estadístico se acerca al valor del parámetro poblacional. Debe buscarse un porcentaje alto de confianza, es decir, una probabilidad alta, ya que una aparente "cercanía" entre el valor calculado en la muestra y el parámetro puede ser ficticio (errores en la selección de la muestra).

En la investigación social existe la convención de aceptar como porcentaje de confianza válido el nivel de significancia del **0.05**, el cual significa que el analista tiene el 95% de seguridad para generalizar sin equivocarse. La escogencia del nivel de significación estadística debe hacerse antes de calcular el porcentaje de probabilidad. Si el resultado estadístico muestra una probabilidad más baja, entonces *la hipótesis nula* deberá ser rechazada. Frecuentemente, en los diseños cuantitativos, los niveles de significancia varían en un rango de **0.10 a 0.001**. Si seleccionamos el nivel de significancia de **0.05**,



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

sólo aquellos resultados que *sean iguales o menores al valor escogido, permitirá rechazar la hipótesis nula*. Únicamente reduciéndolo, se podrá reducir la probabilidad de rechazar la hipótesis nula cuando deberíamos aceptarla.

- Otros métodos para prueba de hipótesis

Todos los test estadísticos tienen básicamente el mismo procedimiento para pruebas concernientes a la diferencia entre una media muestral y una media poblacional; la diferencia entre dos medias muestrales, entre una proporción poblacional y entre dos proporciones muestrales. Las variantes dependen del grado de conocimiento que se tenga sobre los parámetros poblacionales, así:

- Prueba t de Students

Cuando en un estudio, de tipo experimental, se quiere comparar dos medias aritméticas, la prueba que se utiliza es la llamada *t de Student*. La hipótesis en juego permite establecer si existen diferencias significativas entre dos grupos respecto a sus medias⁵⁰. El nivel de medición de las variables debe ser de tipo intervalos, razón o proporcional. Para determinar si el valor "t" es significativo, hay que utilizar la Tabla de Distribución *t de Student*. Se trata de comparar el valor calculado con el valor de la Tabla, basándonos en el nivel de confianza elegido.

- Ejemplificación

A un grupo de diez sujetos se les pidió que leyera un texto sencillo y a otro grupo de diez se le pidió que leyera un texto complejo. Se predijo que se recordarían más palabras del texto sencillo que del texto complejo.

Tabla 6.5. De palabras recordadas de un texto sencillo y de un texto complejo

Grupo 1 (texto sencillo)	Grupo 2 (texto complejo)
<i>Puntajes</i>	<i>Puntajes</i>
10	2
5	1
6	7
3	4
9	4

⁵⁰ La prueba "t" se puede utilizar para comparar resultados de un pretest con los resultados de un post-test.

	8		5
	7		2
	5		5
	6		3
	5		4
Media₁:	6.4	Media₂:	3.7

Hipótesis de trabajo: ¿Existe diferencia significativa entre las medias obtenidas por ambos grupos? – *el promedio de 6.4 palabras recordadas del Grupo 1 es estadísticamente diferente al promedio de 3.7 palabras recordadas del Grupo 2*- La pregunta de investigación hacia referencia a sí era posible afirmar que las diferencias mostradas en la capacidad de recordación eran significativamente validas.

Esquemáticamente la contrastación de hipótesis se presenta así:

H₀: *No existen* diferencias estadísticamente significativas de puntajes entre los sujetos del grupo 1 (que leyeron un texto sencillo) y los del grupo 2 (texto complejo).

H₀: Puntaje Grupo 1 = Puntaje Grupo 2

H_i: *existen diferencias* estadísticamente significativas de puntaje entre los sujetos del grupo 1 (que leyeron un texto sencillo) y los del grupo 2 (texto complejo).

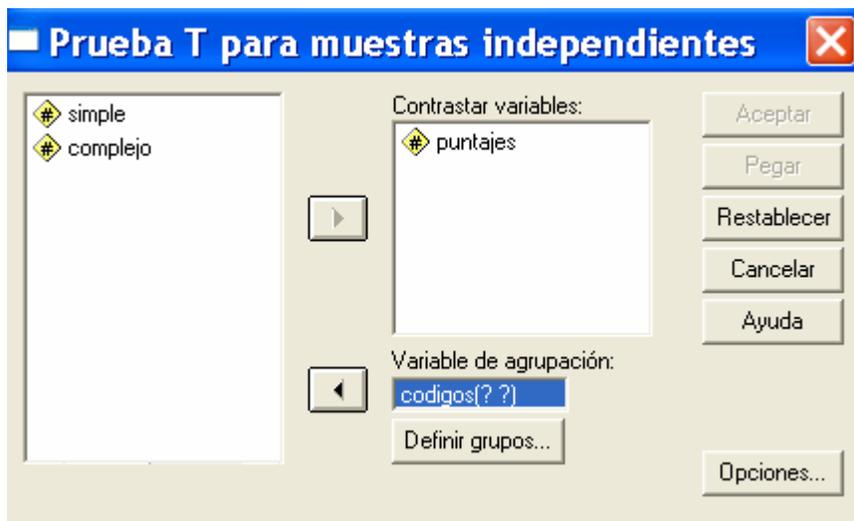
H_i: Puntaje Grupo 1 es diferente al Puntaje Grupo 2

- Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos »»»»»»»»»» *Abrir* »»»»»»»» *Datos*
- Seleccionar el archivo de trabajo **recuerdos.sav** en el CD-ROM.
- Crear una *1ª variable de comparación: puntajes* compuesta por los valores obtenidos en el grupo 1 (*texto sencillo*) y grupo 2 (*texto complejo*)
- Crear una *2ª variable de comparación: Códigos* con los siguientes valores consecutivos: 1 = sencillo; 2 = complejo

recuerdos - Editor de datos SPSS						
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?						
1 : sencillo		10				
	sencillo	complejo	puntaje	codigo	var	var
1	10	2	10	1		
2	5	1	5	1		
3	6	7	6	1		
4	3	4	3	1		
5	9	4	9	1		
6	8	5	8	1		
7	7	2	7	1		
8	5	5	5	1		
9	6	3	6	1		
10	5	4	5	1		
11	.	.	2	2		
12	.	.	1	2		
13	.	.	7	2		
14	.	.	4	2		
15	.	.	4	2		
16	.	.	5	2		
17	.	.	2	2		
18	.	.	5	2		
19	.	.	3	2		
20	.	.	4	2		

- Seleccionar el menú Analizar »»»»»»» elegir la opción Comparar Medias
- Elegir la prueba t para muestras independientes
- A continuación, aparecerá el cuadro de diálogo Prueba T para muestras independientes
- En primer lugar hay que seleccionar la variable dependiente: *puntajes* y transferirla a la ventana **Contrastar variables**:
- En segundo lugar, seleccionar la variable independiente: *códigos* – en donde están los grupos- y transferirla a la ventana “Variable de agrupación”



- Dar un clic en el botón **Definir grupos**
- A continuación aparecerá el cuadro de diálogo Definir grupos
- Contrastar variables: »»»»»»» *Puntaje*
- Variable de agrupación: »»»»»»» *Código (??)*
- Definir grupos: Como se van a comparar los grupos 1 –sencillo- y 2 –complejo- se escribe 1 en la Casilla al lado derecho del Grupo 1:
- Igualmente se escribe 2 en la casilla al lado derecho del Grupo 2:

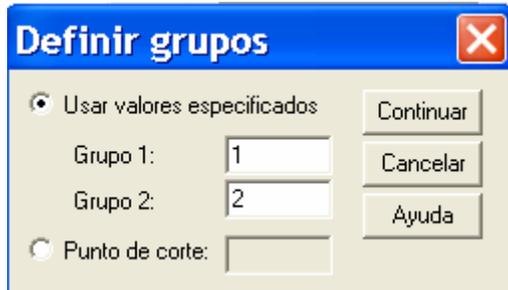


UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

ceo@catios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775



A continuación, en el visor de resultados aparecerá la siguiente tabla:

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia	
									Inferior	Superior
PUNTAJE	Se han asumido varianzas iguales	,423	,524	3,095	18	,006	2,70	,87	,87	4,53
	No se han asumido varianzas iguales			3,095	17,438	,006	2,70	,87	,86	4,54

Para aceptar la H_0 , el valor observado del nivel de significancia de "t"-Sig (bilateral) tiene que ser *mayor o igual* que los valores escogidos - **10% (0.10) ó 5% (0.05)**- Para rechazar la H_0 , el valor observado tiene que **ser menor** al *valor crítico escogido* -se acepta la H_1 -.

Si se elige un nivel de significación del **5% (0.05)** y **el nivel de significación bilateral obtenido es mayor**, se debería aceptar *la hipótesis nula* según la cual "no existen *diferencias estadísticamente significativas* entre el grupo 1 y el grupo 2 y **si es menor** al valor crítico se rechaza la hipótesis de igualdad de puntajes -se acepta la H_1 -.

- Análisis de los datos

A dos grupos de personas se les pidió que leyeran, a los primeros, un texto sencillo y a los segundos, un texto complejo. Se suponía que recordarían más palabras los lectores del texto sencillo que los del texto complejo. Los primeros resultados así lo confirman: el promedio de palabras recordadas por los lectores del texto sencillo -*grupo 1*- fue de 6.4. El promedio de palabras recordadas por los lectores del texto complejo - *grupo 2*, fue de apenas 3.7.

El valor de la prueba t se expresa así **3,095 p ≤ 0,006**. Al analizar el valor del nivel de significación (p Valor) en la Prueba t para la igualdad de medias que es de **0.006**, es decir, *es menor al valor crítico elegido 0,05*, luego hay que aceptar que las diferencias son significativamente válidas. Es decir, la diferencia encontrada en los puntajes de ambos grupos de lectura es válida.

- Comparación de medidas ordinales

Cuando los datos a comparar en dos sub-grupos corresponden a medidas de *nivel ordinal*, como por ejemplo, puntajes en escalas de actitudes y se esta lejos de cumplir con las exigencias de la prueba “t de Student”, se utiliza la *prueba no-paramétrica de Mann-Whitney*, llamada también de *Wilcoxon*, en la forma que se indica en el ejercicio siguiente.

- Ejemplificación

Se ha aplicado una prueba de “conformismo al aprendizaje” a una muestra de adolescentes de los cuales 15 pertenecen a una escuela urbana y 21 a una escuela rural.

Puntajes de aceptación al aprendizaje

Estudiantes Urbanos		Estudiantes Rurales	
Escolar	Puntaje ⁵¹	Escolar	Puntaje
1	12	1	16
2	15	2	22
3	15	3	22
4	40	4	10
5	32	5	12
6	33	6	36
7	22	7	33
8	20	8	27
9	22	9	30
10	21	10	33
11	18	11	38
12	41	12	10
13	45	13	11
14	44	14	17
15	46	15	24
		16	39

⁵¹ El grado de conformismo y aceptación a los modelos pedagógicos de aprendizaje se midió por un puntaje ascendente de 1 a 50 (donde 1 es el mínimo y 50 es el máximo).



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

17	42
18	40
19	26
20	25
21	24

	urbano	rural	puntaje	codigo
1	12	16	12	1
2	15	22	15	1
3	15	22	15	1
4	40	10	40	1
5	32	12	32	1
6	33	36	33	1
7	22	33	22	1
8	20	27	20	1
9	22	30	22	1
10	21	33	21	1
11	18	36	18	1
12	41	10	41	1
13	45	11	45	1
14	44	17	44	1
15	46	24	46	1

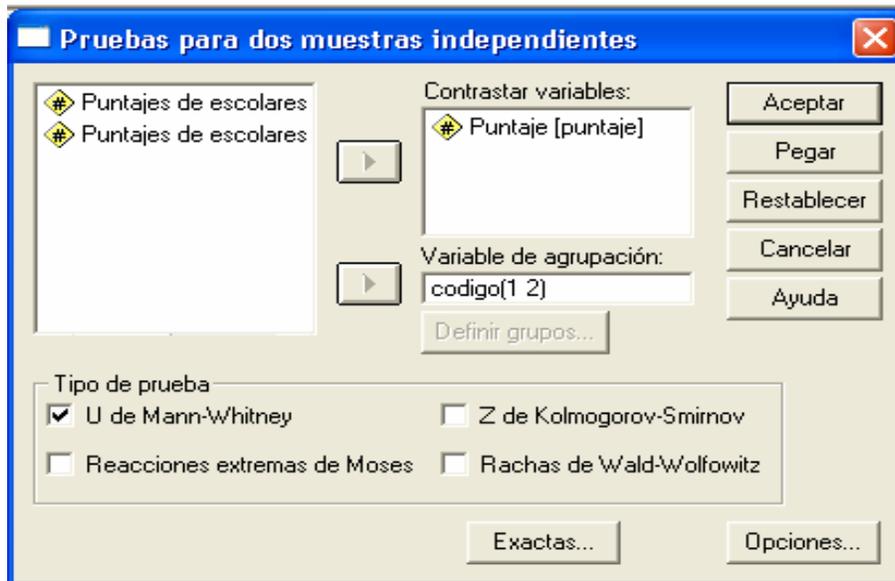
- Seleccionar el menú Analizar »»»»»»»»»»elegir la opción **Pruebas no-paramétricas**
- Seleccionar: **2 muestras independientes Mann-Whitney**⁵²
- A continuación aparecerá el cuadro de diálogo: **Prueba para dos muestras Independientes**
- En primer lugar hay que seleccionar la variable dependiente: *Puntajes* y transferirla a la ventana *Contrastar variables*
- En segundo lugar, seleccionar la variable independiente: *Códigos* – en donde están los grupos- y transferirla a la ventana *Variable de agrupación*

⁵² Se utiliza para comparar dos muestras no relacionadas.



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

- En la sección Tipo de prueba: Seleccionar U de Mann-Whitney



- Dar un clic en el botón **Definir grupos**
- A continuación aparecerá el cuadro de diálogo Definir grupos
- Contrastar variables: »»» Puntaje
- Variable de agrupación: *Código (??)*
- Definir grupos: Como se van a comparar los grupos 1 –urbano- y 2 –rural- se escribe 1 en la Casilla al lado derecho del Grupo 1:
- Igualmente se escribe 2 en la casilla al lado derecho del Grupo 2:
- Definir grupos: Grupo 1 = 1; Grupo 2 = 2

Dos muestras independientes: Definir ... 

Grupo 1:

Grupo 2:

A continuación, en el visor de resultados aparecerá la siguiente tabla:

Prueba de Mann-Whitney

Rangos

COD	N	Rango promedio	Suma de rangos
PUNTAJE Urbano	15	19,67	295,00
Rural	21	17,67	371,00
Total	36		

Estadísticos de contraste^b

	PUNTAJE
U de Mann-Whitney	140,000
W de Wilcoxon	371,000
Z	-,562
Sig. asintót. (bilateral)	,574
Sig. exacta [2*(Sig. unilateral)]	,590 ^a

- Análisis de los datos

El valor de la prueba U de Mann-Whitney se expresa así: $140,000 \geq 0,574$ es decir, es mayor al valor crítico elegido de un 5%, luego hay que aceptar la H_1 que las diferencias son significativas entre los estudiantes urbanos y los rurales.

- Nota metodológica

En nuestro caso, se eligió un nivel de significación de **5%** (0.05) y **el nivel de significación obtenido (Asintótica o exacta) es mayor**, se debería aceptar *la hipótesis*

alternativa según la cual “existen diferencias estadísticamente significativas entre el grupo de puntaje entre los estudiantes de escuelas de nivel urbano y los de nivel rural.

$H_0 = A$ no se diferencia significativamente de B

$H_i = A$ se diferencia significativamente de B

3.1.5. Medidas de relaciones entre variables

En esta Capítulo examinaremos la relación entre variables. De esta forma completaremos todo el escenario que se refiere a la existencia o no de relaciones entre variables medidas en cualquier nivel de medición. De acuerdo a Briones *“Una de las principales funciones de la investigación social -cualquiera que sea el campo disciplinario particular en la cual se efectúa- consiste en buscar relaciones entre el objeto principal de estudio y otros objetos o, si se quiere, entre características principales del objeto de estudio y otras características del mismo o del contexto en el cual se encuentra”* (Briones, Guillermo, 1988:76).

Cuando se habla de posibles relaciones entre dos variables, lo que se esta haciendo es analizar sí: *Existe o no relación entre las variables, la forma o dirección de la relación y la fuerza de la relación.* También Briones señala que *se acostumbra distinguir con el término "asociación" la relación que se da entre características cualitativas o semi-cualitativas (variables nominales y ordinales) y con el término "correlación" la relación entre características cuantitativas.* (Briones, 1988; 76).

- Nota metodológica

Cuando el tipo de medición utilizado es cuantitativo, en el caso específico de variables tipo interval y razón o continuas, el rango de valores varía entre -1.00, lo que significa la existencia de una asociación perfecta de tipo negativa, hasta +1.00, que señala una asociación positiva perfecta. Tales mediciones se dicen ser "direccionales" y reflejan la propiedad de direccionalidad. Con variables cualitativas, a partir del *nivel de significancia*, se podrá determinar la existencia o no de la asociación entre las variables.

Cuadro sinóptico de medidas de relaciones simples entre variables.

Principales medidas de relación de variables nominales, ordinales, intervalo y proporcionales.

	Datos Nominales	Datos ordinales	Datos Intervalos	Datos Proporcionales
Medidas de relación	comparación de porcentajes	tau b		Coeficiente de Pearson
	Coeficientes de Contingencia	Tau c		
	Chi - Cuadrado	D de Somer		
	Coeficiente de Goodman-Kruskal	Gamma		
	Lambda			
			Coeficiente Rho de Spearman	

3.1.5.1. Medidas de asociación entre variables cualitativas.

Examinemos algunas *estadísticas no-paramétricas* que permiten medir asociaciones entre variables medidas a nivel nominal y ordinal.

Prueba del Chi Cuadrado de Pearson: Es una prueba estadística no-paramétrica para **evaluar hipótesis** acerca de la relación entre dos variables cualitativas. Es una prueba que parte del supuesto de **no-relación entre variables** –independencia- y el investigador evalúa si en su caso **esto es cierto o no y esto** es a menudo de interés. Esta es una prueba de significancia estadística no-paramétrica, utilizada con variables medidas a nivel



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

nominal u ordinal que permite determinar si existe o no, una relación-asociación sistemática entre dos variables.

- Ejemplificación

La modalidad de estudio es independiente al programa académico en cual esta vinculado el estudiante. Operacionalmente hablando, esta afirmación se puede plantear por ejemplo, en la siguiente pregunta **¿las variables *Tipo de estudiante según Programa (tipo) y Modalidad de estudio (estudio)* son independientes entre sí?**

Para contestar esta pregunta se hace uso de **la prueba de hipótesis basada en el Chi - Cuadrado**: La **aceptación de H_0** (hipótesis nula) indica **la independencia** entre las variables. Sí se rechaza **H_0** , se acepta **la hipótesis alternativa H_1** , o sea, **la no-independencia (relación)** entre las variables.

La prueba de hipótesis se plantea como:

H_0 : **Independencia** entre las variables - **Tipo de estudiante y modalidad de estudio-**

H_1 : **No-independencia** entre las variables -**Tipo de Estudiante y modalidad de estudio-**

Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos »»»»»»»»»» Abrir »»»»»»»»»» Datos
- Seleccionar el archivo de trabajo **habitos2003.sav** en el CD-ROM
- Seleccionar el menú Analizar »»»»»»»»»» Seleccionar la opción **<Estadísticos descriptivos>** »»»»»»»»»» Activar opción **<Tablas de contingencia>**

A continuación aparecerá el cuadro de diálogo Tablas de contingencia:



- Transferir a la ventanilla Filas, la variable independiente *tipo*
- Transferir a la ventanilla Columnas, la variable dependiente *estudio*
- Dar clic en el botón *Estadísticos...*
- A continuación aparecerá el cuadro de diálogo *Tablas de contingencia: Estadísticos*
- Complete los campos requeridos: marcando el estadístico Chi-cuadrado
- Dar clic en el botón **Continuar**

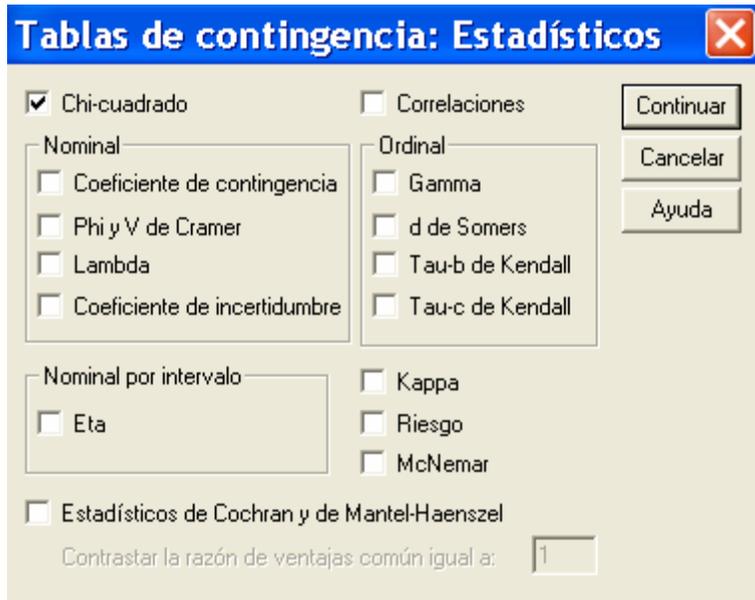


UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

ceo@catios.udea.edu.co

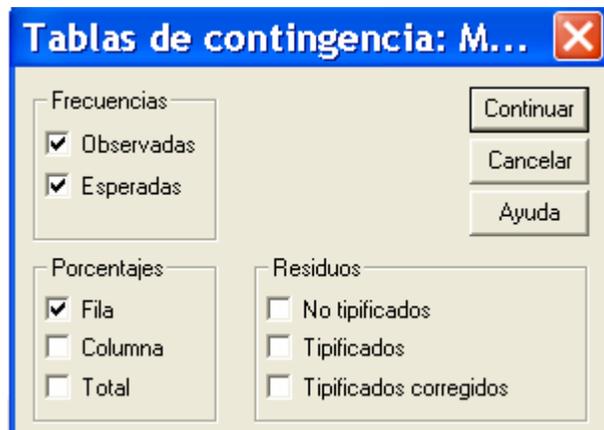
<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775



De un clic sobre el botón **Continuar**

A continuación aparecerá el cuadro de diálogo siguiente:





UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

ceo@catios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

Al completar los campos de acuerdo al modelo anterior: Marcar la opción “Esperadas”
A continuación aparecerá en el Visor de resultados la siguiente tabla:

Tabla de contingencia tipo * estudio

			estudio			Total
			estudio sistemático diario	periodos irregulares de estudio	combinación de los dos tipos	
tipo	ciencias sociales y humanas	Recuento	6	15	23	44
		% de tipo	13.6%	34.1%	52.3%	
	Medicina Veterinaria	Recuento	10	8	24	42
		% de tipo	23.8%	19.0%	57.1%	
Total		Recuento	16	23	47	86
		% de tipo	18.6%	26.7%	54.7%	

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	3.107 ^a	2	.212
Razón de verosimilitud	3.150	2	.207
Asociación lineal por lineal	.099	1	.753
N de casos válidos	86		

- Análisis de los datos

La pregunta que se planteó es ¿existe relación entre el tipo de estudiante según programa académico y la modalidad de estudio que utilizan? El contraste de Chi-cuadrado comprueba que si existe independencia entre ambas variables. El valor del Chi cuadrado de Pearson de **3.107** $p \leq 0,212$ indica que “el tipo de estudiante” y “la modalidad de estudio” son independientes ya que la probabilidad de que una muestra aleatoria arroje un valor de Chi-cuadrado de al menos dicha magnitud es de **0.212**. Esta probabilidad también se conoce como **nivel de significación observado** de la prueba.

Si la probabilidad es suficientemente pequeña (habitualmente menor que **0.05 ó 0.10**), se rechazará la hipótesis de que *las dos variables son independientes*. Pero, es mayor, de acuerdo a los resultados es **0.212**, se debe aceptar que *no hay relación o asociación entre las variables*.

Por lo tanto, se puede aceptar la hipótesis nula, de la independencia⁵³ entre la modalidad de estudio y el tipo de estudiantes (de las facultades de Ciencias Sociales y de Medicina Veterinaria).

- Nota metodológica

En síntesis, el Chi-cuadrado permite decidir si las variables, de tipo cualitativo, son independientes o asociadas y, si es que existe no indica que tan fuerte es esa relación.

Conjuntamente con Chi-cuadrado existen otras medidas para evaluar si las variables incluidas en una tabulación están asociadas, veamos algunas de ellas:

- Comparación de porcentajes: Es una medida de asociación que consiste en comparar los porcentajes en una tabla, en el cual una de las variables es considerada como causal o independiente y sirve como base del cálculo de los porcentajes.
- Lambda: Es una medida de asociación para cruces de variables que se usa comúnmente con variables medidas a nivel nominal.

⁵³ No existe asociación entre las variables.

3.1.5.2. Asociación entre variables nominales

En esta Capítulo se introduce la noción de relación entre variables en contraste con los diseños experimentales que predicen diferencias entre las condiciones experimentales. Es decir, un investigador podría estar interesado no tanto en el desempeño de los dos grupos preseleccionados, sino más bien en observar cómo todo el conjunto del rango de los puntajes en una variable, por ejemplo **los puntajes de capacidad lectora**, se relaciona con todo el conjunto del rango de puntajes en otra variable, por ejemplo los puntajes que se representan **puntajes en estadística** obtenidos por los estudiantes de la Facultad de Administración de Empresas.

En un caso como éste un investigador mediría los puntajes de las dos variables en un solo grupo de sujetos, con el fin de investigar hasta qué punto las características individuales de cada uno en cada variable se correlacionan con su desempeño en la otra variable. La pregunta por resolver es si la habilidad en lectura y la habilidad en estadísticas “*van juntas*” (*asociadas*), o si no hay ninguna conexión entre los dos tipos de habilidades.

- Ejemplificación

Supongamos que a un investigador le interesa saber si existe relación entre la habilidad lectora y la habilidad en estadística. El objetivo sería más bien estudiar si los estudiantes que obtienen un puntaje alto o bajo en una variable, también obtienen puntajes altos o bajos en la otra.

Tabla 6.6. Puntajes en estadística y en lectura

Estudiante	Puntaje de Estadística	Puntaje de lectura
1	9	2
2	10	3
3	12	1
4	6	1
5	11	4
6	9	1
7	12	5
8	16	8
9	13	5
10	10	3
11	13	6
12	14	7

Hipótesis de trabajo:

H₀: No existe asociación entre los puntajes obtenidos en estadística y los puntajes obtenidos en habilidad lectora *-independencia entre las variables-*

H_i: Existe asociación entre los puntajes obtenidos en estadística y los de habilidad en lectura.

- **Nota metodológica:** Para datos nominales (clasificatorios) se utilizan los siguientes coeficientes: *coeficiente de contingencia, coeficiente Phi, coeficiente V. de Cramer, coeficiente Lambda y coeficiente de incertidumbre.* Finalmente, digamos aquí que cuando se trata de muestras, la posible asociación debe ser sometida a una prueba de significación estadística.

- Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos »»»»»»»» Abrir »»»»»»»»»»»» Datos
- Seleccionar el archivo de trabajo **nominal.sav** en el CD-ROM
- En la barra de Menús, seleccionar el menú Analizar
- Elija la opción **<Estadísticos descriptivos>** »»»»»»»» **<Tablas de contingencia>**
- En el cuadro de diálogo **Tablas de contingencia** »»»»»»»» Estadísticos:
- Seleccionar los estadísticos para **variables nominales**:
- *Coefficiente de contingencia C; Phi y V. de Cramer.*
- En el cuadro de diálogo *Tablas de contingencia: mostrar en las casillas:*
Porcentajes: *seleccionar filas*
- A continuación, en el visor de resultados aparecerá la siguiente tabla:

Medidas simétricas			
		Valor	Sig. aproximada
Nominal por nominal	Phi	2,273	,101
	V de Cramer	,859	,101
	Coefficiente de contingencia	,915	,101
N de casos válidos		12	

a. Asumiendo la hipótesis alternativa.
b. Empleando el error típico asintótico basado en la hipótesis nula.

- Análisis de los datos

Como el nivel de significación es mayor al 5% -valor crítico escogido-, se acepta la H_0 , es decir, que no existe relación entre la habilidad lectora y los puntajes obtenidos en habilidad estadística.



- Nota metodológica

H_0 = no existe asociación entre las variables »»»»»» sí es mayor al valor crítico escogido

H_1 = existe asociación entre las variables »»»»»» sí es menor al valor crítico escogido

3.1.5.3. Asociación entre variables ordinales: Coeficiente de asociación por rangos de Spearman

Determina la asociación entre dos variables medidas en escalas ordinales. Una forma de medir el grado de asociación entre dos variables medidas a nivel ordinal la proporciona el coeficiente de asociación de Spearman o de rangos.

- Ejemplificación

Suponga que nos interesa averiguar si hay alguna relación entre la participación de los estudiantes en deportes al aire libre -medida sobre una escala desde 1 (nunca juega) hasta 10 (siempre juega) y la frecuencia con que los mismos estudiantes se ausentan – medida en una escala desde 1 (con frecuencia – hasta 5 - rara vez).

Tabla: Puntajes de participación de estudiantes en deportes y frecuencia de ausentismo.

Sujeto	Variable A	Variable B
	Puntaje en escala de participación deportiva (1 – 10)	Puntaje en escala de frecuencia de ausentismo (1 – 5)
1	5	2
2	2	3
3	3	7
4	10	5
5	9	4
6	9	5
7	2	4
8	6	3
9	3	1
10	4	1



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

11
12

8
10

4
5

Hipótesis de trabajo: La hipótesis de trabajo afirma que una alta participación de los estudiantes en las actividades deportivas esta relacionada con bajas tasas de ausentismo académico.

H₀: No existe asociación entre los puntajes obtenidos de participación deportiva y ausentismo académico (Independencia entre las variables).

H_i: existe asociación entre los puntajes en participación deportiva y ausentismo académico.

	Puntaje en escala de participación deportiva (1 – 10)	<u>Rango A</u>⁵⁴	Puntaje en escala de frecuencia de ausentismo (1 – 5)	Rango B
Sujeto				
1	5	5.0	2	3.5
2	3	2.5	2	3.5
3	7	7.0	4	7.5
4	10	12	5	11
5	9	9.5	4	7.5
6	9	9.5	5	11
7	2	1.0	4	7.5
8	6	6.0	3	5.0
9	3	2.5	1	1.5
10	4	4.0	1	1.5
11	8	8.0	4	7.5
12	10	12	5	11

⁵⁴ Crear rangos para ambas variables



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

significativo, debe ser *igual o mayor* que los valores críticos escogidos - **0.10 ó del 0.05** o, lo que es lo mismo adoptar el nivel de significación del 10% o del 5%.

ceo@carios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

A continuación, en el visor de resultados aparecerá la siguiente tabla:

Correlaciones no paramétricas

Correlaciones

			RANGO A	RANGO B
Rho de Spearman	RANGO A	Coeficiente de correlación	1.000	.780**
		Sig. (bilateral)	.	.003
		N	12	12
	RANGO B	Coeficiente de correlación	.780**	1.000
		Sig. (bilateral)	.003	.
		N	12	12

** La correlación es significativa al nivel 0,01 (bilateral).

- **Análisis de los datos:** el nivel de significación estadística encontrado es de *0.003*, es decir, menor al valor crítico escogido (0.05), por lo tanto, no existe asociación entre los puntajes que miden la participación deportiva y el ausentismo académico.

3.1.5.4. Análisis de regresión

El análisis de regresión y correlación permite mostrar cómo deberíamos determinar tanto la naturaleza como la fuerza de la relación que se puede existir entre dos variables. También es posible predecir el valor de una variable desconocida basados en observaciones pasadas de esa u otra variable.

Lo anterior exige, en muchos casos, el uso de una técnica que permita usar toda la información lograda referente a un juego de variables y al mismo tiempo, describir la relación entre dos o más variables en una forma simple y precisa; ante tal requerimiento, el análisis de regresión es la técnica más apropiada.

- Regresión lineal

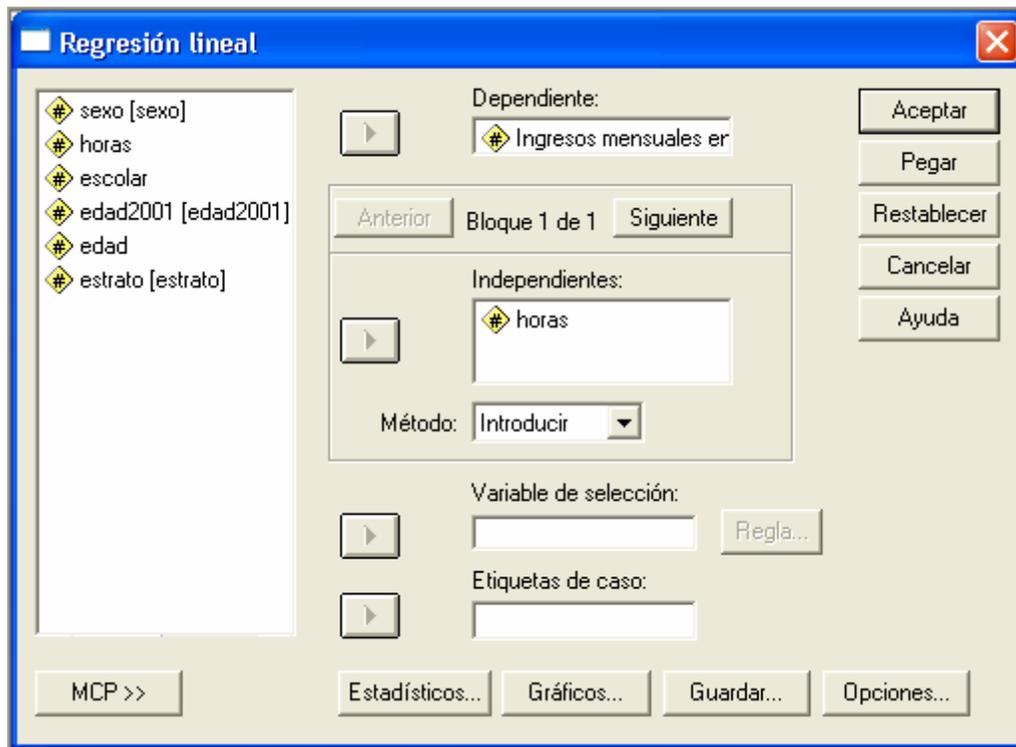
Esta es una técnica estadística que permite analizar la relación entre una variable dependiente – criterio- y un a o más variables – predictoras- El nivel de medición de las variables tiene que ser de intervalo o razón. A su vez, las variables del estudio deben ser, una considerada como independiente y la otra como dependiente -decisión de tipo teórico-metodológica-.

Permite resumir los datos y cuantificar la naturaleza y fuerza de la relación entre variables. También se puede usar ésta técnica para **predecir** o estimar nuevos valores de la variable dependiente de acuerdo a los valores observados en la variable independiente.

- Ejemplificación

Supongamos que necesitamos analizar el impacto de un programa de capacitación en sistematización en el nivel de ingresos de un grupo de egresados del Servicio Nacional

- Seleccionar y transferir la variable *ingresos mensuales* a la ventanilla *Dependiente*.
- Seleccionar la o las variables predictoras o independientes –horas de capacitación- a la ventanilla *Independiente* y dar clic en el botón **Aceptar**



- Nota metodológica

El primer valor es *la variable dependiente ingresos mensuales percibidos* y se trasfiere a la casilla *Dependiente*: -el segundo, es *la variable independiente horas* - número de horas de capacitación en sistematización- y se trasfiere a la casilla *Independientes*:

A continuación, en el visor de resultados aparecerá la siguiente tabla:

Variables introducidas/eliminadas ^b			
Modelo	Variables introducidas	Variables eliminadas	Método
1	HORAS ^a		Introducir

a. Todas las variables solicitadas introducidas
b. Variable dependiente: Ingresos mensuales en dólares

Resumen del modelo				
Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,948 ^a	,898	,893	132,05

a. Variables predictoras: (Constante), HORAS

- Análisis de los datos

Lo primero que interesa conocer es la fuerza de la correlación entre la variable predictora *horas* (número de horas de capacitación) y la variable criterio *Ingresos* (ingresos percibidos). Esto se mide con el *coeficiente de correlación R cuadrado*- $R^2 = .898$, ello indica que el 89,8 por ciento de la varianza de la variable ingresos está predicha por la variable horas de capacitación.

- Nota metodológica

El coeficiente R^2 puede variar entre 0 y 1. Cuando $R^2 = 1$ quiere decir que el 100% de la variabilidad de la variable criterio es explicada por la variable predictora. Cuando $R^2 = 0$ significa que no hay ninguna relación lineal entre la *variable predictora* y la *variable criterio*.



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

ceo@catios.udea.edu.co

<http://ceo.udea.edu.co>

Ciudad Universitaria Bloque 9-252 Telefax: 2105775

3.1.5.5. Coeficientes de correlación

A menudo, los analistas sociales están interesados en descubrir cuál o cuáles variables están más estrechamente asociadas a una determinada variable dependiente, en este caso los coeficientes de correlación se convierten en la herramienta estadística adecuada. En síntesis, el análisis de correlación es una herramienta que se puede usar para describir el grado en el que una variable está relacionada con otra.

Hasta ahora sólo ha interesado saber si hay o no una relación entre variables (si es significativa o no) pero nada se sabe si la relación obtenida es fuerte o débil -fuerza de la relación-. Cuando existe la relación, es conveniente preguntarse: ¿Cuál es su fuerza? Analicemos las formas más utilizadas para detectar la fuerza de la relación entre variables

- Correlación de Pearson

El coeficiente de correlación de Pearson r mide básicamente la naturaleza de la relación entre dos variables cuantitativas. No solo sirve para indicar si hay o no relación, sino también para indicar *la fuerza –grado- y dirección* de la misma. Consiste en una prueba "paramétrica" que mide el grado de correlación entre *los puntajes* obtenidos en dos variables e indica el nivel de significación de la correlación observada. La prueba de Pearson requiere de una medición de los datos en una escala interval o de razón.

El coeficiente de correlación lineal r refleja el grado de relación entre dos variables y lo expresa con un número que va desde "-1" (correlación negativa perfecta) pasa por "0" (no existe correlación) hasta **+1** (correlación positiva perfecta). En síntesis, puede tomar valores comprendidos entre -1 y 1. Si el coeficiente es positivo expresa *una covariación* de las variables en el mismo sentido, y un coeficiente negativo, lo contrario.

La fuerza o intensidad de la relación puede ser débil o fuerte o inclusive puede no existir. Generalmente se acepta el criterio según la cual, si la relación entre las variables se acerca a cero (0), indica debilidad y si se acerca a uno 1 sería indicativo de una alta asociación o correlación.

- Ejemplificación

Mediante una prueba se ha medido *la habilidad⁵⁶ en comprensión y lectura* y el uso de las herramientas estadísticas en un grupo de estudiantes del curso de Introducción a la Investigación social.

Se ha hipotetizado que existe una correlación positiva entre la habilidad en comprensión y lectura de textos y la habilidad en la utilización de la herramienta estadística, es decir, que los estudiantes que tienen un buen desempeño en comprensión y lectura de textos lo tendrán también en el uso de la estadística, como herramienta de análisis.

Tabla 6.8. Puntajes obtenidos en la prueba de lectura y comprensión y estadística social.

Alumno	Variable A ⁵⁷	Variable B ⁵⁸
1	30	20
2	50	10
3	60	20
4	73	12
5	87	16
6	24	7

⁵⁶ Instrumentos de medición generan puntajes cuantitativos que miden las capacidades de uso de las herramientas.

⁵⁷ Variable A: puntaje obtenido en prueba que miden el grado de utilización de la herramienta estadística como elemento de análisis.

⁵⁸ Variable B: puntaje obtenido en prueba de comprensión y lectura de textos.



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

7	44	9
8	35	18
9	97	19
10	62	13
11	49	6
12	93	10

Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos »»»»»»»»»» Abrir »»»»»»»»»» Datos
- Seleccionar el archivo de trabajo **nominal.sav** en el CD-ROM
- En la barra de Menús, seleccionar el menú Analizar »»»»»»»» Correlaciones
- Seleccionar »»» Bivariadas »»»» el cuadro de diálogo: *Correlaciones bivariadas*.
- Seleccionar y transferir la *variable independiente* y la *variable dependiente* a la ventanilla **Variables**
- En la sección Coeficientes de correlación elegir **Pearson**
- En la sección Prueba de significación elegir **Bilateral**
- Al presionar el botón **Aceptar** aparecen los resultados en el Visor

Resultados en el Visor:

Correlaciones

		puntaje en estadística	Puntaje en lectura
puntaje en estadística	Correlación de Pearson	1,000	,857**
	Sig. (bilateral)	.	,000
	N	12	12
Puntaje en lectura	Correlación de Pearson	,857**	1,000
	Sig. (bilateral)	,000	.
	N	12	12

** . La correlación es significativa al nivel 0,01 (bilateral).

- Análisis de los datos



UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS
CENTRO DE ESTUDIOS DE OPINIÓN

Se puede concluir que existe una correlación significativa entre las dos variables: Habilidad en estadística social y habilidad en comprensión y lectura, ya que el coeficiente de correlación entre ellas sí es significativo – $r = 0,857$ $p \leq 0,000$

- Nota metodológica

En el gráfico siguiente se observa una posición relativa de un coeficiente r :

-1.00	-0.70	-0.50	-0.20	0.0	0.20	0.50	0.70	+1.00
Correlación negativa alta	negativa fuerte	moderada	débil o baja		moderada	fuerte	Correlación positiva alta	

- Algunos comentarios sobre el uso de la regresión y correlación

Una pregunta básica tiene que ver con la oportunidad del uso de las técnicas de regresión. Más adelante, en este Capítulo, cuando hablemos de **regresión múltiple**, volveremos sobre este interrogante. Por lo pronto, digamos que raramente un analista social trabaja en el estudio del impacto de una sola variable. El foco de atención generalmente se centra en el impacto o determinación de dos o más variables independientes sobre un hecho o proceso social. En casos como estos, la técnica más adecuada es la de regresión múltiple.

Acerca de la conveniencia de usar las técnicas de correlación o las de análisis de regresión, es evidente que esta última tiene mucho más que ofrecer para un analista social. Fundamentalmente porque motiva al investigador a considerar más cuidadosamente la dirección causal entre las variables, por ejemplo, se deberá tener claro cuál es la variable independiente y cuál es la dependiente. También, el resultado del cálculo de una regresión: el coeficiente de regresión tendrá en términos generales, un significado más sustantivo que un coeficiente de correlación.

3.1.6. Análisis multivariado

Existen algunas técnicas estadísticas que han sido desarrolladas para estudiar problemas complejos desde una perspectiva multivariada, es decir, considerar en el análisis, todas las posibles interrelaciones entre las variables, para lograr un mayor entendimiento y explicación de un proceso. En síntesis, el análisis multivariado es aquel que analiza la relación entre varias variables independientes y, al menos una variable dependiente.

3.1.6.1. Regresión múltiple

En lecciones anteriores describimos las técnicas de relación entre variables. La discusión se centró en el análisis de dos variables, por ejemplo, el efecto de una variable independiente en una dependiente. Sin embargo, la utilidad del modelo de regresión es mejor percibido a través del uso de la técnica de regresión múltiple. Esta herramienta estadística permite analizar la naturaleza de una variable dependiente a partir de 2 o más variables independientes o descriptoras. Sirve para predecir el valor de una variable dependiente conociendo el valor y la influencia de las variables independientes incluidas en el modelo.

Permite evaluar el valor individual de cada variable independiente en la variable dependiente mientras se controla tal influencia por el impacto de otras variables de tipo cuantitativas al igual que la influencia de todas las variables independientes combinadas.

Otra de las ventajas de la regresión múltiple sobre el análisis de tablas de continencia se basa en la cantidad de variables de control o intervinientes que se pueden utilizar; en las tablas de continencia, se limita el uso a una o dos variables intervinientes; en cambio, con la regresión múltiple no existe tal limitación.

- Ejemplificación

Se aplicó una encuesta a los estudiantes destinada a conocer la opinión sobre sus cursos y el desempeño de los profesores que los dictan. Se les pidió que evaluaran tanto al profesor como al curso de acuerdo a la siguiente escala de calificaciones:

Muy deficiente 1 deficiente 2 aceptable 3 bueno 4 excelente 5

Objetivo: Estudiar la relación entre la evaluación del profesor de un curso realizada por sus estudiantes y cinco características del profesor y del curso: Dominio de los temas del curso, empleo de recursos didácticos, respeto en el trato con los estudiantes, evaluación de ejercicios y exámenes, evaluación del interés de los contenidos.

- Procedimiento SPSS

- En la barra de Menús, elija el menú Archivos »»»»»»»»»» Abrir »»»»»»»»»» Datos
- Seleccionar el archivo de trabajo **encuesta al estudiante sobre el curso y el profesor.sav** en el CD-ROM.
- En la barra de Menús, seleccionar el menú Analizar »»»»» Regresión »»»»» Lineal
- Seleccionar y transferir las variables predictoras o independientes: Dominio de los temas, empleo de recursos didácticos, respeto en el trato con estudiantes, evaluación de ejercicios y exámenes y evaluación del interés de los contenidos a la ventanilla Independientes: y dar un clic sobre el botón Aceptar
- Dar un clic sobre el botón Estadísticos
- A continuación aparecerá el cuadro de diálogo Regresión lineal: Estadísticos
- Seleccionar y marcar las opciones: Estimaciones; Descriptivos y Correlaciones parcial y semiparcial
- Dar un clic sobre el botón **Continuar**

Regresión lineal: Estadísticos

Coeficientes de regresión

- Estimaciones
- Intervalos de confianza
- Matriz de covarianza

Ajuste del modelo

- Cambio en R cuadrado
- Descriptivos
- Correlaciones parcial y semiparcial
- Diagnósticos de colinealidad

Residuos

- Durbin-Watson
- Diagnósticos por caso
 - Valores atípicos a más de desviaciones típicas
 - Todos los casos

A continuación, en el visor de resultados aparecerá la siguiente tabla:

Estadísticos descriptivos

	Media	Desviación típ.	N
Calificación del Docente de 1 a 5	3,84	,55	25
Dominio de los temas	3,92	,49	25
Empleo recursos didácticos	4,24	,72	25
Respeto en el trato con estudiantes	3,92	,49	25
Evaluación ejercicios y exámenes	3,80	,65	25
Evaluación del interés de los contenidos	3,84	,69	25

En la tabla *Estadísticos descriptivos* se muestran los promedios de la evaluación estudiantil y la desviación estándar o típica de cada una de las seis variables, lo que permite visualizar los promedios de evaluaciones y los casos extremos.

Correlaciones

	Calificación del Docente de 1 a 5	Dominio de los temas	Empleo recursos didácticos	Respeto en el trato con estudiantes	Evaluación ejercicios y exámenes	Evaluación del interés de los contenidos
Calificación del Docente de 1 a 5	1,000	,714	,516	,561	,839	,477
Dominio de los temas	,714	1,000	,640	,486	,733	,329
Empleo recursos didácticos	,516	,640	1,000	,640	,642	,583
Respeto en el trato con estudiantes	,561	,486	,640	1,000	,733	,697
Evaluación ejercicios y exámenes	,839	,733	,642	,733	1,000	,488
Evaluación del interés de los contenidos	,477	,329	,583	,697	,488	1,000

- Análisis de los datos

El peso o influencia que tiene cada variable independiente sobre la variable dependiente se muestra en la tabla correlaciones. En términos generales, se puede afirmar que hay algunas correlaciones significativas entre las variables predictoras o independientes – *dominio del tema y evaluación de ejercicios y exámenes* con la evaluación del profesor.

A continuación en el Visor de resultados aparecen los coeficientes que miden la correlación de todas las variables independientes en la variable dependiente – *calificación del profesor*-

Resumen del modelo				
Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,874 ^a	,763	,701	,30

a. Variables predictoras: (Constante), Evaluación del interés de los contenidos, Dominio de los temas, Respeto en el trato con estudiantes, Empleo recursos didácticos, Evaluación ejercicios y exámenes

ANOVA ^b						
Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	5,617	5	1,123	12,249	,000 ^a
	Residual	1,743	19	9,172E-02		
	Total	7,360	24			

a. Variables predictoras: (Constante), Evaluación del interés de los contenidos, Dominio de los temas, Respeto en el trato con estudiantes, Empleo recursos didácticos, Evaluación ejercicios y exámenes

Un $R^2 = 0.763$ indica que el 76 % de la variación total de los promedios de evaluación se explica por estas 5 variables que se usaron. Se puede concluir que el R^2 es significativo, ya que F es igual a **12,249** $p \leq 0.000$

4. Conclusión

En los capítulos anteriores se ha descrito en qué consiste la medición y la recolección de datos, los diversos tipos de instrumentos que se utilizan con este objetivo, y como se codifica y sistematiza. Finalmente, en este Capítulo se han presentado y ejemplificado las técnicas estadísticas más utilizadas, en el análisis de los datos, tales como las técnicas paramétricas y las no-paramétricas, haciendo énfasis en la lógica y oportunidad de su uso. Implícitamente lo que se pretende responder es la pregunta metodológica básica:

¿Cuál estadístico se debe usar y qué significa su resultado?

Hoy en día, el cálculo de los estadísticos es realizado por programas informáticos especializados de manejo y gestión de datos. Y, por ello, en este texto se da mayor importancia a *la lógica y oportunidad del uso de estos estadísticos* en relación con los objetivos y con el tipo de información recolectado. Lo importante es saber qué técnicas de análisis se deben utilizar cuando se desea describir un colectivo o una situación, encontrar una relación entre unas variables o, hacer algunas inferencias estadísticas.

BIBLIOGRAFÍA

AIGNEREN, Miguel. (1997), *INVESTIGACIÓN CUANTITATIVA EN CIENCIAS SOCIALES. PROCESAMIENTO Y ANÁLISIS DE DATOS*. Universidad de Antioquia, Facultad de Ciencias Sociales y Humanas, Centro Estudios de Opinión, Medellín, Colombia.

----- (1999), *DISEÑO Y ANÁLISIS DE LA INFORMACIÓN CUANTIFICABLE EN LA INVESTIGACIÓN SOCIAL*, Universidad de Antioquia, Facultad de Ciencias Sociales y Humanas, Centro Estudios de Opinión, Medellín, Colombia.

-----, (1991), *EL ANÁLISIS DEL DATO TABULADO*, Universidad de Antioquia. Facultad de Ciencias Sociales y Humanas, Departamento de Sociología, Medellín. Colombia.

Briones, Guillermo, (1988), *MÉTODOS Y TÉCNICAS DE INVESTIGACIÓN APLICADA A LA EDUCACIÓN Y A LAS CIENCIAS SOCIALES, MÓDULO 2, TIPOS DE INVESTIGACIONES Y DISEÑOS METODOLÓGICOS*, Corporación Editorial Universitaria, Bogotá, Colombia.

-----, (1986), *CURSO AVANZADO DE TÉCNICAS DE INVESTIGACIÓN SOCIAL APLICADAS A LA EDUCACIÓN, MÓDULO 2, TÉCNICAS E INSTRUMENTOS PARA LA RECOCAPÍTULO DE INFORMACIÓN*, PIIE, Santiago de Chile.

----- (1988), *MÉTODOS Y TÉCNICAS AVANZADAS DE INVESTIGACIÓN APLICADAS A LA EDUCACIÓN Y A LAS CIENCIAS SOCIALES, MÓDULO 1*, PIIE. Santiago de Chile.

----- (1988), *MÉTODOS Y TÉCNICAS AVANZADAS DE INVESTIGACIÓN APLICADAS A LA EDUCACIÓN Y A LAS CIENCIAS SOCIALES*, MÓDULO 4, ANÁLISIS E INTERPRETACIÓN, PIIE, Santiago de Chile, ICFES, Bogotá, Colombia.

-----, (1990), *MÉTODOS Y TÉCNICAS DE INVESTIGACIÓN PARA LAS CIENCIAS SOCIALES*, Editorial Trillas, México.

Barbbie, Earl, (1979), *THE PRACTICE OF SOCIAL RESEARCH*, Wadsworth Publishing Co. Belmont, California.

Blalock, Hubert, Jr., Blalock B. Ann, (1978), *METHODOLOGY IN SOCIAL RESEARCH*, McGraw-Hill Co., Nueva York.

Escalante, Carlos, (1983), *LA MEDICIÓN DE LAS ACTITUDES*, Ediciones Tercer Mundo. Bogotá, Colombia.

-----, (1987), *EL PROBLEMA Y LA HIPÓTESIS*, MÓDULO 2, ICFES, Bogotá, Colombia.

Facundo D., Ángel, (1987), *EL PROYECTO DE INVESTIGACIÓN, MÓDULO 5, SERIE APRENDER A INVESTIGAR*, ICFES, Bogotá, Colombia

Ghiglione, Rodolfo, et al, (1989), *LAS ENCUESTAS SOCIOLÓGICAS, TEORÍA Y PRÁCTICA*. Editorial Trillas, México,

Hernández, S. Roberto y et. Al, (1991), *METODOLOGÍA DE LA INVESTIGACIÓN*, Editorial McGraw-Hill, México.

Hitchens, Christopher, (1992), *ENCUESTAS: ¿LAS MALAS DEL PASEO?*, El Tiempo, Suplemento Dominical, Bogotá, Colombia.

Mayor, Alberto, (1987), *LA RECOCAPÍTULO DE LA INFORMACIÓN*. SERIE APRENDER A INVESTIGAR, ICFES, Procesos editoriales, Bogotá, Colombia.

Noelle, Elizabeth, (1970) *ENCUESTAS EN LA SOCIEDAD DE MASAS*, Alianza Editorial, Madrid.

Puerta Zapata Antonio, (1996), *EVALUACIÓN Y SISTEMATIZACIÓN DE PROYECTOS SOCIALES: UNA METODOLOGÍA DE LA INVESTIGACIÓN*. Universidad de Antioquia, Facultad de Ciencias Sociales, Medellín Colombia.

Rodríguez, Humberto, (1987) *EL INFORME DE INVESTIGACIÓN*, MÓDULO 6, SERIE APRENDER A INVESTIGAR, ICFES, Bogotá, Colombia.

Ruiz, Jaime; Serna, Alba Lucía; Aignerren, Miguel, (1991) *EL MEDELLÍN QUE YO QUIERO*, Universidad de Antioquia, Facultad de ciencias Sociales. Centro de Estudios de Opinión, Concejo de Medellín. Medellín, Colombia.

Ruiz. Jaime, Aignerren, Miguel, et, al. (1991), *IMÁGENES DE IDENTIDAD, INTEGRACIÓN Y CONFLICTO ENTRE LA COMUNIDAD ESTUDIANTIL DE LA UNIVERSIDAD DE ANTIOQUIA*, Universidad de Antioquia, Centro de Estudios de Opinión, Medellín, Colombia.

Serna A., Alba Lucía, et al., (1992), *LAS TÉCNICAS DE RECOPIADO DE INFORMACIÓN*, MÓDULO 3, ICFES – INER, Procesos Editoriales ICFES, Bogotá, Colombia.

Siegel, Sydney, (1978), *ESTADÍSTICA NO-PARAMÉTRICA APLICADA A LAS CIENCIAS DE LA CONDUCTA*", Trillas, México.

Vélez B., Eduardo, (1987), *EL ANÁLISIS DE LA INFORMACIÓN*, MÓDULO 4, SERIE APRENDER A INVESTIGAR, ICFES, Bogotá, Colombia.

Webb, Kenneth y Hatry, Harry, (1973), *OBTAINING CITIZEN FEEDBACK: THE APPLICATION OF CITIZEN SURVEYS TO LOCAL GOVERNMENT*, The Urban Institute, Washington.

GLOSARIO DE TÉRMINOS

Actitud. Reacción o predisposición favorable o desfavorable hacia algo o alguien, que se manifiesta en las creencias o en un comportamiento proyectado.

Aleatorio. Suceso o hecho producido al azar.

Análisis de Regresión. Técnica estadística que establece una relación funcional entre una variable dependiente y una serie de variables dependientes. Si solo hay una variable independiente, se denomina análisis simple; si hay varias es múltiple.

Archivo. Conjunto de registros o datos relacionados entre sí, los cuales se pueden almacenar en una base de datos.

Categoría. Es una medida de una variable en la que la escala utilizada es de tipo nominal u ordinal.

Codificación de los Datos. Consiste en asignar códigos numéricos a las respuestas dadas a un cuestionario para poder efectuar el tratamiento estadístico de los datos.

Código. Convención para representar letras, números y caracteres especiales con el fin de identificar características o categorías de respuestas.

Coeficiente Alfa de Cronbach. Coeficiente que estima la consistencia interna de una escala de medición.

Comportamiento organizacional. Estudio de los individuos y de los grupos dentro del ámbito de la organización.

Curtosis. Grado en que las observaciones están agrupadas en los extremos o colas de la distribución normal.

Confiabilidad. Grado en que una prueba proporciona resultados consistentes.

Contraste de Hipótesis. Conjunto de procedimientos estadísticos que permiten decidir si una hipótesis se debe aceptar o rechazar.

Datos. Valores alfanuméricos que procesa el computador para convertirlos en información. Se entiende por **dato** un valor específico de una variable. Por ejemplo, 30 años es un dato de la variable edad.

Dígito. Cada uno de los números que integran un dato numérico; por ejemplo: 2003 tiene cuatro dígitos, o sea, cuatro números.

Digitación. Es el registro de los códigos de las variables en un programa especializado para su posterior sistematización y análisis, por ejemplo, digitación de los datos en la matriz de datos del SPSS.

Diagrama de causa y efecto. Diagrama usado para organizar y mostrar, en figuras o gráficos, las posibles causas de un problema.

Efecto de Halo. Situación que se da cuando el entrevistador permite que una característica u opinión particular contamine la opinión sobre los demás rasgos-respuestas que se esperan.

Encuesta Piloto. Aplicación de un cuestionario de prueba a una submuestra con el objetivo de identificar problemas con las preguntas y su nivel de respuesta.

Encuesta de Panel. Muestra permanente de unidades de análisis de los cuales se obtiene información periódicamente.

Estándar. Unidad de medida adoptada y aceptada comúnmente como criterio para realizar algún proceso.

Estimación estadística. Procedimiento estadístico que trata de determinar las características de una población a partir de las características obtenidas de una muestra representativa.

Grabación de Datos. Almacenamiento de los datos en un medio magnético.

Grupo de Control. Grupo de unidades de análisis o sujetos que en un estudio experimental no son expuestos a ninguno de los tratamientos cuyos efectos tratan de medirse.

Inferencia Estadística. Proceso estadístico inductivo que trata de obtener información sobre las características de una población a partir de los datos obtenidos en una muestra.

Muestra Intencional. Muestra de unidades de análisis o individuos elegidos mediante un muestreo no probabilística. Se trata de elegir aquellas unidades de análisis que de acuerdo al investigador resuman las características deseadas o a su alcance.

Submuestra. Conjunto de unidades de análisis –sujetos u objetos- obtenidos de una muestra, de acuerdo a un criterio determinado, para efectuar análisis específicos en ella.

Tabla de Contingencia. Tabla de doble entrada en la que se presentan los valores de las frecuencias conjuntas de dos o más variables.

Tabulación Cruzada. Técnica estadística que permite calcular la distribución de las frecuencias de una variable para cada una de las categorías de la otra variable.

Tipo de Variable. En el programa SPSS, hay que definir si la variable es numérica, continua, discreta, clasificatoria, etc.

Validación de los datos. Es la inspección de las respuestas de los cuestionarios con el fin de asegurar que estén suficientemente contestados y que las respuestas sean consistentes.

Validez. El grado en que una prueba mide lo que se pretende que mida.

Variable. En relación con SPSS es la característica que varía que se mide u observa de un elemento, una muestra o la población; ejemplos: edad, escolaridad, estado civil, etc.

Variable Criterio. Variable dependiente o explicada por la variable independiente o explicatoria.

Varianza. Fluctuación promedio de un determinado valor de la población dada.

Unidad de Análisis. Es el sujeto – objeto del estudio - por ejemplo, la aplicación de un cuestionario a un grupo de 205 estudiantes implica la existencia de 205 casos o unidades de análisis.

ANEXOS

Archivos de trabajo

Se adjuntan 10 archivos de trabajo:

1. Conformismo.sav
2. Deportes.sav
3. Encuesta al estudiante y al profesor.sav
4. Habitos.sav
5. Ingresos.sav
6. Nominal.sav
7. Profesional-2003.sav
8. Recuerdos.sav
9. Situación.sav
10. Temor ciudadano.sav