

REVISTA SIGMA

Departamento de Matemáticas

Universidad de Nariño

Volumen XV^N1 (2019), páginas 28–44

TECNICAS ESTADISTICAS EN EL ANÁLISIS CUANTITATIVO DE DATOS

Arsenio Hidalgo Troya ¹

Abstract: In the process of knowledge construction, in the field of social research, there is generally talk of qualitative and quantitative designs. The quantitative design is a systematic approach to the study of a reality based mainly on the analysis of categorical and numerical variables and preferably using statistical techniques that allow to show the trend of the data and establish, through functional relationships, the interrelation and dependence on variables for explanatory and predictive purposes.

Keywords.

Descriptive and exploratory analysis, statistical inference, parametric and non-parametric techniques, multivariate analysis.

Resumen: En el proceso de construcción del conocimiento, en el campo de la investigación social, generalmente se habla de diseños cualitativos y cuantitativos. El diseño cuantitativo es una forma de aproximación sistemática al estudio de una realidad apoyándose principalmente en el análisis de variables categóricas y numéricas y utilizando preferentemente técnicas estadísticas que permiten mostrar la tendencia de los datos y establecer, a través de relaciones funcionales, la interrelación y la dependencia de las variables con fines explicativos y predictivos.

Palabras Clave. Análisis descriptivo y exploratorio, inferencia estadística, técnicas paramétricas y no paramétricas, análisis multivariado.

1. Introducción

La investigación de tipo cuantitativo está enfocada bajo el *pensamiento positivista*. A. Comte (1798-1857) afirmaba que el espíritu humano debe renunciar a conocer el ser mismo (la naturaleza) de las cosas y contentarse con las verdades que proporciona la observación y la experimentación. Esta corriente relega la subjetividad humana y busca verificar empíricamente los hechos y sus causas, con el claro objetivo de establecer leyes universales. La complejidad de todo lo humano se reduce a variables que, cuantificadas y analizadas, facilitan el cálculo de la probabilidad estadística de que ocurra un evento. El positivismo, precisamente, es el origen del desarrollo formal del *método científico* y de la *investigación*

¹Departamento de Matemáticas y Estadística Centro de Estudios y Asesorías en Estadística-CEASE Universidad de Nariño. Email: arsenio.hidalgo@gmail.com

cuantitativa, que con frecuencia, se la asimila como la única forma válida alcanzar el conocimiento.

Como reacción al positivismo, nacen corrientes filosóficas, como la *fenomenología* impulsada por Husserl (1859-1938), que fundamentan la *investigación cualitativa*, planteando que la realidad se puede conocer a través de la abstracción teórica, analizando las cualidades de la experiencia, que permiten entender la esencia misma del fenómeno observado.

Es muy importante valorar la naturaleza del problema-pregunta a estudiar, pues esta consideración marca la elección del camino a seguir, es decir la utilización de una metodología cuantitativa o cualitativa. Algunas características epistemológicas que marcan la diferencia entre estos dos enfoques se presentan en la tabla 1.

<i>Característica</i>	<i>Cuantitativa</i>	<i>Cualitativa</i>
<i>Percepción de la realidad</i>	Objetiva, excluyente	Subjetiva, incluyente
<i>Razonamiento</i>	Deductivo, contrasta hipótesis	Inductivo, genera hipótesis
<i>Finalidad</i>	Comprobación, confirmación, reducción	Exploración, descubrimiento, expansión
<i>Orientación</i>	Al resultado	Al proceso
<i>Principio de verdad</i>	Particularista, estable (permanente), predeterminada, centrada en similitudes	Holística, dinámica (provisoria), se construye, centrada en diferencias
<i>Perspectiva del investigador</i>	Desde fuera (al margen de los datos)	Desde dentro (próximo a los datos)
<i>Causalidad</i>	Antecedente específico	Interacción de factores

Tabla 1. Características epistemológicas de la investigación cuantitativa y cualitativa

La investigación cuantitativa permite a las ciencias naturales operar según un modelo determinista de causas y efectos, un modelo que se aplica a los seres humanos lo mismo que a las plantas y seres inanimados. En general aceptamos que dicho modelo es el adecuado en tales casos. En las ciencias sociales la postura determinista contrasta con la idea del *libre albedrío*, el cual lo damos por hecho en nuestra vida diaria, estas operan sobre la base de un modelo causal *probabilístico*. Los conocimientos que buscamos cuando analizamos los datos de las investigaciones sociales comprenden de manera inevitable un modelo determinista de la conducta humana. Al contemplar las razones de que las personas sean como son, se asume implícitamente que sus características y acciones están determinadas por fuerzas y factores que operan con ellas. No tenemos que creer que estamos totalmente determinados ni tenemos que vivir como si lo fuéramos, pero debemos estar dispuestos a aplicar la lógica determinista en la búsqueda de explicaciones cuando realicemos investigación en las ciencias sociales.

En las ciencias sociales los conceptos son los significados aceptados que asignamos a los términos para facilitar la comunicación, la medición y la investigación. Una *definición conceptual* define un término (variable conceptual) con otro(s) término(s). Esta definición es importante pero insuficiente para medir la variable de investigación, porque no nos relacionan directamente con la realidad, lo que realmente se tiene es una definición nominal. Como señala Kerlinger “. . .los científicos debe ir más allá. Deben definir las variables que se usan en sus hipótesis en forma tal que puedan ser comprobadas, Esto es posible usando lo que se conoce como definiciones operacionales” [11, p.41]. Una *definición operacional* constituye el conjunto de procedimientos que describe las actividades que un observador debe realizar para obtener las impresiones sensoriales, las cuales indican la existencia de un concepto teórico en mayor o menor grado, es decir especifica en que actividades u operaciones deben realizarse para medir una variable de investigación.

El siguiente diagrama la progresión de las etapas de medición desde la idea vaga de lo que significa el término hasta la medición o mediciones específicas de una investigación científica completamente estructurada:

Conceptualización Definición nominal Definición operacional Medición real

La medición o mediciones obtenidas en la última etapa del diagrama anterior permiten obtener un conjunto de datos de la(s) variable(s) que hacen parte de la investigación y que conforman una estructura de información llamada comúnmente la *base de datos*. Esta estructura informática se puede construir de manera electrónica utilizando software computacional (hoja de cálculo, paquetes estadísticos u otras herramientas) y ser procesada utilizando técnicas estadísticas apropiadas para su análisis e interpretación de resultados, teniendo en cuenta los objetivos planteados y las hipótesis formuladas en la investigación.

Este capítulo pretende abordar, de manera sintetizada, las técnicas que permiten analizar los datos e interpretar sus resultados en investigaciones con diseño cuantitativo.

2. PROCESAMIENTO CUANTITATIVO DE LOS DATOS

Como se mencionó anteriormente, la Estadística es la herramienta fundamental para hacer análisis a los datos obtenidos en una investigación de tipo cuantitativa. Para apoyar el procesamiento de datos estadísticos, existen en el mercado informático una gran variedad de software estadístico tales como IBM-SPSS, SAS, StatGraphics, Stata, Epi Info, R, etc., con sus ventajas y desventajas, como son el precio, la facilidad de manejo y su funcionalidad. En el campo de las ciencias sociales se puede considerar como muy apropiado el primero de los softwares mencionados SPSS (Statistical Package for the Social Sciences), 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 archivo de los datos obtenidos y el análisis estadístico de los mismos.

El uso de estos paquetes estadísticos facilita enormemente el procesamiento de bases de datos en una investigación, por lo que se hace necesario conocer muy bien cuáles son los procedimientos o técnicas estadísticas más apropiadas para alcanzar los objetivos de la investigación y verificar las hipótesis que se hayan formulado o que están implícitas en el trabajo de investigación.

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, los estudiantes deben familiarizarse con 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. Se deben abordar los aspectos operativos relativos al diseño de archivos de trabajo mediante los diferentes módulos de programa estadísticos como SPSS.

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.

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:

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.

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.

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.

La aplicación de los diferentes métodos estadísticos depende de varios factores a considerar en el proceso de investigación como son los objetivos, las hipótesis, la naturaleza de las variables, las propiedades de los datos, el tipo de investigación y los alcances de la misma, entre otros.

Así se habla de *métodos estadísticos descriptivos*, cuando se pretende mirar la tendencia de los datos y describir las características de las variables que intervienen en la investigación. De *métodos estadísticos inferenciales*, cuando su objeto es extrapolar o generalizar a una población de estudio, los resultados obtenidos en una muestra probabilística seleccionada de la misma. Se aplican *pruebas paramétricas*, cuando los datos cumplen las propiedades de independencia, normalidad y homogeneidad y *pruebas no paramétricas* cuando los datos no satisfacen alguna de las dos últimas propiedades mencionadas. Se tiene también diferentes

tipos de análisis según el número de variables que se consideran en el mismo; así el *análisis univariado* permite describir o caracterizar de manera individual el comportamiento de cada variable que interviene en la investigación, el *análisis bivariado* estudia la asociación entre dos variables y el *análisis multivariado* la asociación entre múltiples variables, mediante el uso de modelos estadísticos y herramientas matemáticas.

El análisis estadístico de los datos en una investigación es una combinación de las técnicas o métodos mencionados en el párrafo anterior.

Es conveniente señalar que un investigador en el campo social debe saber qué tipo de estadísticos son más pertinentes y cómo leer o analizar los resultados. La importancia de lo anterior se basa en la idea que el investigador no debe usar a ciegas unos estadísticos, es decir, debe ser capaz de juzgar su pertinencia en relación con el problema de estudio, y leer y evaluar la calidad de los resultados. Generalmente, un investigador ciego considera que la objetividad de los resultados están avalados por el carácter que toma todo resultado estadístico cuando es producto casi maravilloso de un proceso de sistematización, lo cual conduce a muchos errores.

A continuación se presenta de manera resumida las técnicas estadísticas descriptivas e inferenciales, paramétricas y no paramétricas y en las dimensiones univariada, bivariada y multivariada más utilizadas y apropiadas según la naturaleza de las variables y los objetivos del análisis de los datos.

3. ANÁLISIS DESCRIPTIVO Y EXPLORATORIO DE LOS DATOS

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 característica 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 se entiende en 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, como se anotó anteriormente, de una serie de elementos, entre los cuales, los más importantes a considerar son:

- Los objetivos del estudio
- La hipótesis de investigación
- La naturaleza o 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, 1988: 23)

Entonces la primera tarea del investigador 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, comparaciones, tipologías o diferencias entre ellos.
- Relaciones entre las variables de estudio.

Las tareas anteriores se realizan aplicando técnicas estadísticas descriptivas, tanto gráficas como numéricas, que se resumen en las tablas **2** y **3**, respectivamente, de acuerdo a la dimensión y naturaleza de las variables.

Sin embargo, mucha de la información contenida en los datos, esencial para realizar un buen análisis estadístico, no queda bien reflejada si, tan sólo, nos quedamos con la que nos aporta la *estadística descriptiva*, de ahí que, en los últimos años, se propugna por conjugar la estadística descriptiva clásica con otro tipo de análisis (donde los gráficos adquieren gran importancia) conocido como *Análisis Exploratorio* y que fue introducido en 1977 por J. W. Tukey, para quien el analista de datos debe ser un profesional cuya actuación es similar a la de un detective.

El análisis exploratorio de los datos, implica una actitud curiosa que está motivada por la premisa de que cuanto mejor conozca el investigador los datos que tiene, más eficientemente se pueden usar para desarrollar y refinar la teoría. Por tanto, el principal objetivo de un abordaje exploratorio de datos es maximizar lo que podemos aprender a partir de ellos.

<i>Dimensión</i>	<i>Naturaleza variables</i>	<i>Técnica</i>
Univariada	Cualitativa	Tablas de frecuencia, barras simples, pastel
	Cuantitativa	Tablas de frecuencia, histograma, polígono de frecuencias, ojiva, tallos y hojas, cajas
Bivariada	Dos Cualitativas	Tablas de contingencia, barras compuestas
	Dos cuantitativas	Dispersión
	Cualitativa y cuantitativa	Cajas
Multivariada	Cuantitativas y cualitativas	Gráficos cartesianos, dispersogramas, rostros de Chernoff, curvas de Andrews

Tabla 2. Técnicas descriptivas gráficos

El análisis exploratorio de datos es previo a la aplicación de pruebas estadísticas paramétricas y de técnicas multivariadas, pues permite verificar las premisas o supuestos que deben cumplir los datos para la aplicación de esas técnicas.

Así una tarea del análisis exploratorio y que debe llevarse a cabo antes de aplicar cualquier técnica multivariante es el análisis de los *datos ausentes* y la detección de *datos atípicos* (en inglés *outliers*) Cualquier recogida y proceso de datos presenta problemas que van a impedir obtener información de algunos de los elementos de la población en estudio, pero la existencia de datos ausentes nunca debe impedir la aplicación del análisis multivariante, ni tampoco debe limitar la posibilidad de generalizar los resultados de una investigación. El analista deberá identificar la presencia de datos ausentes y llevar a cabo las acciones necesarias para intentar minimizar sus efectos. Igualmente se debe detectar la existencia de observaciones que no siguen el mismo comportamiento que el resto, es decir casos atípicos, éstos suelen deberse a errores en el procedimiento a la hora de introducir los datos o de codificarlos. También pueden ser consecuencia de acontecimientos anormales que hacen destacar determinadas observaciones. Una vez detectados los casos atípicos el analista debe saber elegir con criterio entre eliminarlos del análisis o evaluar toda la información incluyéndolos.

<i>Dimensión</i>	<i>Naturaleza variables</i>	<i>Técnica</i>
Univariada	Cualitativa	Proporciones, razones y tasas
	Cuantitativa	Medidas de Localización: media, mediana, moda, percentiles
		Medidas de Dispersión: rango, varianza, desviación típica, coeficiente de variación
		Medidas de forma: coeficientes de asimetría y curtosis
Bivariada	Dos Cualitativas	Coeficientes de asociación
	Dos cuantitativas	Análisis de correlación y regresión lineal
	Cualitativa y cuantitativa	Coeficientes asociación, Análisis de discriminante y regresión logística
	Multivariada	Cuantitativas y cualitativas

Tabla 3. Técnicas descriptivas numéricas

Otra tarea importante que realiza el análisis exploratorio de datos es la comprobación de los supuestos subyacentes en los métodos paramétricos y multivariantes. Estos supuestos suelen ser el contraste de la *normalidad* de todas y cada una de las variables que forman parte del estudio, el testeo de la linealidad de las relaciones entre las variables que intervienen en el estudio (la relación entre la posible variable dependiente y las variables independientes que la explican ha de ser una ecuación lineal), la comprobación de la *homoscedasticidad* de los datos que consiste en ver que la variación de la variable dependiente que se intenta explicar a través de las variables independientes no se concentra en un pequeño grupo de valores independientes (se tratará por tanto de ver la igualdad de varianzas para los datos agrupados según valores similares de la variable dependiente) y la comprobación de la *multicolínealidad* o existencia de relaciones entre las variables independientes. A veces también es necesario contrastar la ausencia de correlación serial de los residuos o *autocorrelación*, que consiste en asegurar que cualquiera de los errores de predicción no está correlacionado con el resto.

Actualmente se utilizan las novedosas técnicas del análisis exploratorio de datos, mediante las cuales se pueden analizar los datos exhaustivamente y detectar las posibles anomalías

que presentan las observaciones. Los estadísticos descriptivos más habitualmente utilizados han sido la media y la desviación típica. Sin embargo, el uso automático de estos índices no es muy aconsejable. La media y la desviación típica son índices convenientes sólo cuando la distribución de datos es aproximadamente normal o, al menos, simétrica y unimodal. Pero las variables objeto de estudio no siempre cumplen estos requisitos. Por lo tanto es necesario un examen a fondo de la estructura de los datos.

Se recomienda iniciar un análisis exploratorio de datos con gráficos que permitan visualizar su estructura. Por ejemplo, para datos cuantitativos es aconsejable comenzar con el gráfico de tallo y hojas o histograma digital. El paso siguiente suele ser examinar la posible presencia de normalidad, simetría y valores atípicos (outliers) en el conjunto de datos. Para ello suelen utilizarse los gráficos de caja y bigote. No obstante los gráficos de caja siempre deben ir acompañados de los histogramas digitales (o gráficos de tallo y hojas), ya que los primeros no detectan la presencia de distribuciones multimodales. Los gráficos de dispersión nos dan una idea de las relaciones entre variables y su ajuste. El uso de estadísticos robustos (o resistentes) es muy aconsejable cuando los datos no se ajustan a una distribución normal. Estos estadísticos son los que se ven poco afectados por valores atípicos. Suelen estar basados en la mediana y en los cuartiles y son de fácil cálculo. Fruto del análisis exploratorio, a veces es necesario transformar las variables.

La mayoría del software estadístico dispone de herramientas que aportan técnicas gráficas preparadas para el examen de los datos que se ven mejoradas con medidas estadísticas más detalladas para su descripción. Estas técnicas permiten examinar las características de la distribución de las variables implicadas en el análisis, las relaciones bivariantes (y multivariantes) entre ellas y el análisis de las diferencias entre grupos. Hay que tener en cuenta que las representaciones gráficas nunca sustituyen a las medidas de diagnóstico formal estadístico (contrastes de ajuste de los datos a una distribución, contrastes de asimetría, contrastes de aleatoriedad, etc.), pero proporcionan una forma alternativa de desarrollar una perspectiva del carácter de los datos y de las interrelaciones que existen, incluso si son multivariantes.

4. ANÁLISIS INFERENCIAL LOS DATOS

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, porque no se recolectan datos de toda la población, pero pueden ser inferidos de los estadígrafos, de ahí el nombre de estadística inferencial.

(Hernández, R. y et. Al, (1991): 373)

La estadística inferencial se utiliza fundamentalmente en dos procesos: *estimación de parámetros poblacionales y pruebas de hipótesis*.

En esta sección, 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 35 % de los drogadictos que han asumido individualmente su terapia (tratamientos tradicionales) reinciden en estados de drogadicción.

Se plantea la hipótesis de que sólo el 20 % 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 de la población de reincidentes. 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 muestra. Luego se juzga si la diferencia es significativa.

Mientras menor sea la diferencia, mayor será la probabilidad de que el valor hipotético esté correcto. Pero, en no todas las ocasiones, la diferencia entre el parámetro hipotético de la población y el estadístico de la 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.

El ejemplo 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.

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.

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 diseñar el *muestreo de ambos grupos de rehabilitación*. Lamentablemente no hay técnica que nos defina una muestra 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.

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 comprobación 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_1).
- Elección del riesgo o nivel de significancia que prácticamente es el complemento de lo que se llama nivel de confianza. Normalmente = 0,05 ó 0,01.
- Contraste de hipótesis. Comparar el valor encontrado del estadístico de prueba obtenido con el valor crítico o el valor p de significación estadística (valor que suministran los paquetes estadísticos) 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 y luego se recogen los datos 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. Hay dos tipos de hipótesis estadísticas: Una, es la hipótesis de investigación o alterna, que señala la posible existencia o relación entre dos hechos. La otra, es la hipótesis nula, que se define para que el investigador evalúe su hipótesis de trabajo o investigación.

Debemos tener 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*, *F de Snedecor (ANOVA)* o *Chi-Cuadrado*, con los cuales el investigador podrá decidir sobre la validez de su propuesta.

El objetivo de una prueba de hipótesis es hacer un juicio sobre la *diferencia* entre el valor de la muestra (estadístico) y el valor hipotético de la población (parámetro). 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, entonces debemos decidir con qué porcentaje de confianza este estadístico se acerca al valor del parámetro. 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 0,05 a 0,001. Si seleccionamos el nivel de significancia de 0,05, 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.

Todos los test estadísticos tienen básicamente el mismo procedimiento para pruebas concernientes a la diferencia con una media poblacional; la diferencia entre dos medias poblacionales, con una proporción poblacional y entre dos proporciones poblacionales, también para comprobar la asociación entre variables. Las variantes dependen del grado de conocimiento que se tenga sobre las propiedades y distribución de los datos, la característica de la prueba. La tabla 4 presenta un resumen de diferentes tests estadísticos para comparación de poblaciones y asociación de variables.

Es conveniente señalar finalmente que, la estimación de parámetros a través de los llamados intervalos de confianza, permite también el contraste de las hipótesis estadísticas.

<i>Distribución</i>	<i>Variable Independiente</i>	<i>Variable Dependiente</i>	<i>Relación entre muestras</i>	<i>Prueba Estadística</i>
Normal (Paramétricos)	Una sola muestra (valor teórico)	Cuantitativa		t-student para una muestra
	Dicotómica (2 muestras)	Cuantitativa	Independientes	t-student muestras independientes
			Relacionadas	t-student muestras relacionadas
	Policotómica (k muestras)	Cuantitativa	Independientes	ANOVA de una vía
			Relacionadas	ANOVA de medidas repetidas
	Cuantitativa	Cuantitativa		t-student
				ANOVA
				Correlación Pearson
No normal (No paramétricos)	Una sola muestra (valor teórico)	Catagórica		Binomial
		Cuantitativa		Test de signos
				Test de Wilcoxon
	Dicotómica (2 muestras)	Catagórica	Independientes	Chi-cuadrado de Pearson
				Test exacto de Fisher
				Chi-cuadrado de Mantel-Haenzel
				Prueba de Kolmogorov-Smirnov
				Prueba de las Rachas
		Catagórica	Relacionadas	Test de McNemar
		Cuantitativa	Independientes	Test U de Mann-Whitney
				Z Kolmogorov-Smirnov
				Rachas de Wald-Wolfowitz
				Reacciones extremas de Moses
		Cuantitativa	Relacionadas	Test de Wilcoxon
	Policotómica (k muestras)	Catagórica	Independientes	Chi-cuadrado de Pearson
		Catagórica	Relacionadas	Prueba Q de Cochran
				W de Kendall (concordancia)
		Cuantitativa	Independientes	Prueba de Kruskal-Wallis
				Test Mediana
				Honckeere Tersptra
			Relacionadas	Prueba de Friedman
	Cuantitativa	Cuantitativa		Correlación de Spearman

Tabla 4. Tipo de test estadístico para hacer inferencias, comparación y asociación.

5. MÉTODOS ESTADISTICOS MULTIVARIADOS

Generalmente en un estudio real, el investigador dispone habitualmente de muchas variables medidas u observadas en un grupo de individuos y pretende estudiarlas conjuntamente, para lo cual suele acudir al análisis estadístico de datos univariado, y bivariado, previo a un análisis multivariado. Entonces se encuentra frente a una diversidad de técnicas y debe seleccionar las más adecuada a sus datos pero, sobre todo, a sus objetivos de investigación.

Al observar muchas variables sobre una muestra es presumible que una parte de la información recogida pueda ser redundante o excesiva, en cuyo caso los métodos multivariados de reducción de la dimensión (análisis en componentes principales, factorial, correspondencias, escalamiento óptimo, homogeneidades, análisis conjunto, etc.) tratan de eliminarla. Estos métodos combinan muchas variables observadas para obtener pocas variables ficticias que las representen.

Por otro lado, las unidades o individuos pueden presentar ciertas características comunes en sus datos o respuestas, que permitan intentar su clasificación en grupos de cierta homogeneidad. Los métodos de clasificación (análisis cluster, análisis discriminante, árboles de decisión, etc.) buscan analizar las relaciones entre variables para ver si se pueden separar los individuos en agrupaciones a posteriori.

Finalmente, podrá existir una variable cuya dependencia de un conjunto de otras sea interesante detectar para analizar su relación o, incluso, hacer su predicción cuando las demás sean conocidas. En este cabe incluir la regresión lineal simple y múltiple, regresión no lineal, regresión logística, análisis de la varianza simple y múltiple, las técnicas de análisis de series temporales, etc.

Por otra parte, el investigador tendrá que considerar si asigna a todas sus variables una importancia equivalente, es decir, si ninguna variable destaca como dependiente principal en el objetivo de la investigación, en ese caso se creó los *métodos multivariados de interdependencia*. Si es así, porque maneja simplemente un conjunto de diversos aspectos observados y coleccionados en su muestra, puede acudir para su tratamiento en bloque a lo que podría llamarse técnicas multivariadas descriptivas.

Puede realizarse con dos orientaciones diferentes: por una parte, para reducir la dimensión de una tabla de datos excesivamente grande por el elevado número de variables que contiene y quedarse con unas cuantas variables ficticias que, aunque no observadas, sean combinación de las reales y sinteticen la mayor parte de la información contenida en sus datos. En este caso también se deberá tener en cuenta el tipo de variables que maneja. Si son variables cuantitativas, las técnicas que le permiten este tratamiento pueden ser el Análisis de componentes principales y el Análisis factorial, si son variables cualitativas, puede acudir al Análisis de correspondencias y si son variables cualitativas ordinales se acude al Escalamiento multidimensional. La Tipología acepta variables cualitativas y cuantitativas.

Por otra parte, la otra orientación posible ante una colección de variables sin ninguna destacada en dependencia, sería la de clasificar sus individuos en grupos más o menos homogéneos en relación al perfil que en aquéllas presenten, en cuyo caso utilizará por ejemplo el Análisis de clusters, donde los grupos, no definidos previamente, serán configurados por las propias variables que utiliza. La tabla 5 resume algunas de estas técnicas teniendo en cuenta la naturaleza de las variables.

<i>Variables</i>	<i>Agrupación</i>	<i>Técnica</i>
Cualitativas	Categorías de variables	Análisis de Correspondencias
Cuantitativas	Variables	Análisis Factorial
Cuantitativas y cualitativas	Objetos	Análisis de Conglomerados
Cuantitativas y cualitativas	Objetos	Escalamiento Multidimensional

Tabla 5. Métodos Multivariados de Interdependencia.

Si no fuera aceptable una importancia equivalente en las variables, porque alguna variable se destaca como dependiente principal, habrá de utilizarse *métodos multivariadas de dependencia*, considerando la variable dependiente como explicada por las demás variables independientes explicativas, y tratando de relacionar todas las variables por medio de una posible ecuación o modelo matemático que las ligue.

El método elegido podría ser entonces la Regresión lineal, generalmente con todas las variables cuantitativas. Una vez configurado el modelo matemático se podrá llegar a predecir el valor de la variable dependiente conocido el perfil de todas las demás. Si la variable dependiente fuera cualitativa dicotómica (1 = sí; 0 = no) podrá usarse como clasificadora, estudiando su relación con el resto de variables de clasificación a través de la Regresión logística. Si la variable dependiente cualitativa observada constatará la asignación de cada individuo a grupos previamente definidos (dos, o más de dos), puede ser utilizada para clasificar nuevos casos en que se desconozca el grupo a que probablemente pertenecen, en cuyo caso estamos ante el Análisis discriminante, que resuelve el problema de asignación en función de un perfil cuantitativo de variables de clasificación.

Si la variable dependiente es cuantitativa y las explicativas son cualitativas estamos ante los modelos del análisis de la varianza, que puede extenderse a los modelos loglineales para el análisis de tablas de contingencia de dimensión elevada. Si la variable dependiente puede ser cualitativa o cuantitativa y las independientes cualitativas, estamos ante el caso de la Análisis conjunto. La tabla 6 resume algunos de estos modelos teniendo en cuenta la naturaleza de las variables dependientes e independientes.

<i>Variables Dependientes</i>	<i>Variables Independientes</i>	<i>Modelo</i>
Una Cualitativa	Cualitativas y cuantitativas	Lineal General
Una Cuantitativa	Cuantitativas	Regresión Lineal Múltiple
Varias Cuantitativas	Cuantitativas	Correlación Canónica
Una Cuantitativa	Cualitativas	ANOVA
Una Cuantitativa	Cualitativas y cuantitativas	ANCOVA
Varias Cuantitativas	Cualitativas	MANOVA
Varias Cuantitativas	Cualitativas y cuantitativas	MANCOVA
Cualitativa	Cuantitativas	Discriminante
		Regresión Logística
		Análisis Conjunto

Tabla 5. Métodos Multivariados de Interdependencia.

6. CONCLUSIONES

Los párrafos anteriores, nos dan una visión técnica y práctica del proceso de investigación social que se apoya en los diseños cuantitativos tratando de acercarnos a la práctica real del investigador. Se insiste en la tendencia de combinar 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.

Este artículo aborda la presentación de las herramientas estadísticas más útiles en el tratamiento cuantitativo de las variables de una investigación, buscando dejar claro una serie de conceptos y técnicas.

Referencias

- [1] 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.
- [2] Hernández, R. y et. Al, (1991), Metodología de la Investigación, Editorial McGraw-Hill, México.
- [3] Aigner, M. (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.
- [4] Barbbie, E. (2000), Fundamentos de la Investigación Social, Internacional Thomson México.
- [5] Blalock, H., Jr., Blalock, A. (1978), Methodology in Social Research, *McGraw-Hill Co.*, Nueva York.
- [6] De la Garza, J., Morales, B., González, B. (2013), Análisis Estadístico Multivariante. Un enfoque teórico y práctico. *Mc Graw Hill*, México.
- [7] Escalante, Carlos, (1983), La Medición de las Actitudes, Ediciones Tercer Mundo. Bogotá, Colombia.
- [8] Escalante, C. (1987), El Problema y la Hipótesis, Módulo 2, ICFES, Bogotá, Colombia.
- [9] Facundo, D (1987), El Proyecto de Investigación, Módulo 5, Serie Aprender a Investigar, ICFES, Bogotá, Colombia
- [10] Hernández, R. y et. Al, (1991), Metodología de la Investigación, Editorial McGraw-Hill, México.
- [11] Kerlinger, F. (1979). Investigación del Comportamiento. *Mc Graw Hill*, Bogotá, Colombia.
- [12] Pérez, C. (2008), Técnicas de Análisis Multivariante de Datos. Aplicaciones con SPSS. *Prentice Hall*, Madrid, España. 30
- [13] Rodríguez, H. (1987) El Informe de Investigación, Módulo 6, Serie Aprender a Investigar, ICFES, Bogotá, Colombia.
- [14] Serna A., Alba L., et al., (1992), Las técnicas de Recolección de la información, Módulo 3, ICFES – INER, Procesos Editoriales ICFES, Bogotá, Colombia.
- [15] Siegel, S. (1978), Estadística No-paramétrica Aplicada a las Ciencias de la Conducta, Trillas, México.

Hidalgo, A.

- [16] Vélez, E. (1987), El Análisis de la Información, Módulo 4, Serie Aprender a Investigar, ICFES, Bogotá, Colombia.