

NOTAS SOBRE EL MANEJO DE LA INFORMACION CUANTITATIVA EN LA INVESTIGACION SOCIAL*

Miguel Aignerren

Abstract

Usually in the process of construction and reconstruction of knowledge in the social sciences we are talking about designs quantitative and qualitative designs, in our particular case, we will refer to a type of design that uses preferably based information expressed in numerical form, as quantitative. The design is a form of quantitative systematic approach to the study of reality that relies mainly into categories and numerical analysis performed through different forms of inter statistically these categories. The emphasis on the analysis of the data, usually relies on the trend shows that the behaviour of them.

Resumen

Generalmente en el proceso de construcción y reconstrucción del conocimiento en el campo de las ciencias sociales estamos hablando de diseños cuantitativos y diseños cualitativos; en nuestro particular caso, nos referiremos a un tipo de diseño que utiliza preferentemente como base una información expresada en forma numérica, como cuantitativo. El diseño cuantitativo es una forma de aproximación sistemática al estudio de la realidad que se apoya preferentemente en categorías numéricas y que realiza el análisis a través de diferentes formas de interrelacionar estadísticamente esas categorías. El énfasis del análisis de los datos, se apoya generalmente en las tendencias que muestra el comportamiento de ellos.

* Plan de maestría en sociología. Universidad del valle. Facultad de ciencias sociales y económicas. Material docente preparado por el prof. M. Aignerren

De acuerdo a la cobertura del estudio, en los diseños cuantitativos se puede hablar de tres tipos de estrategias de aproximación y manejo de la información; estos son:

Estudios de casos. Es el tipo de investigación apropiado para estudiar intensivamente y con profundidad características básicas de grupos reducidos tales como comunidades, grupos de presión, grupos étnicos, profesionales, etc. Es el típico caso de una estrategia investigativa que en la práctica investigativa real es muy difícil de inscribirla en una categoría única, ya que en ella coexisten con mayor o menor predominancia diferentes tipos de investigación y adicionalmente facilita la complementación de instrumentos de recolección de información (desde el cuestionario, la entrevista hasta la historia de vida). El manejar un número reducido de unidades de análisis o sujetos de estudio facilita la aplicación de diferentes modalidades de análisis de información, como son el análisis de contenido del lenguaje escrito o verbal, formas de cotidianidad, rituales sociales de integración o de conflicto, etc. El manejo de este tipo de información tradicionalmente percibida como "cualitativa" no quita que las "calidades apreheñidazas" en el trabajo, puedan ser sistematizadas y codificadas para los efectos de una mejor lectura e interpretación (no se trata tampoco de reivindicar aquí la "manía codificadora").

Estudios censales. Fundamentalmente lo que los define es una metodología de recolección de información extensiva a toda la población.

Estudios muestrales. En este tipo de investigación lo característico es la selección de una "muestra representativa" de sujetos mediante procedimientos estadísticos al azar, de tal forma que se cumpla el requisito de representatividad que facilite la generalización.

En resumen, podemos decir que no hay ventajas absoluta ni desventaja absoluta en ninguno de los sistemas descritos de recolección y manejo de la información. Lo básico es tener en

cuenta sus ventajas y limitaciones de cada uno y seleccionarlos según la problemática y los objetivos del estudio.

Todo análisis de una información debe comenzar con las "descripción y clasificación" de los procesos u hechos que representan, de donde surge la necesidad de adaptar los instrumentos metodológicos al genero o tipo de hechos o procesos que se investigan. Hoy en día, es claro que toda disciplina abarca una dimensión específica de la realidad, lo cual a su vez se deriva de la imposibilidad que tiene el hombre de conocer el conjunto de la realidad o, que para lograr mayor conocimiento de la realidad, sea indispensable una provisión mayor de información; este crecimiento de la cantidad de información aumenta las dificultades para su "análisis" o lectura, a no ser del uso de instrumentos auxiliares como puede ser la estadística_.

Intentaremos vincular los usos que se hacen de estas herramientas, como es la estadística, en la investigación social con las concepciones teóricas vigentes en el campo de las ciencias sociales Lamentablemente con demasiada frecuencia la utilización que se hace de la misma acarrea vicios y errores; estos pueden ubicarse en dos planos:

- *Concepciones empiricistas*. La utilización acrática o la absolutización de las técnicas estadísticas, que pretende sustituir con "la formalización de la lógica matemática", la carencia de elementos teóricos que permitan una real lectura de lo que expresan los "datos" (Hiperfactualismo o sobre valorización de los aspectos objetivos del análisis).

- *Concepciones impresionistas* - subjetivistas: Sobre valorización de lo subjetivo, negando el uso de técnicas estadísticas en el marco de una sobre valorización del análisis cualitativo, producto de concepciones subjetivistas.

Hablemos un poco sobre el papel del instrumento estadístico en la investigación social: Los métodos estadísticos son unos valiosos instrumento auxiliar de conocimiento; por medio de la metodología estadística se logra tanto la sistematización formal de la metodología investigativa que utilice preferentemente datos cuantificables (por ejemplo, la definición de la población muestral), como la derivar conclusiones a partir de observaciones que no son, en si, absolutamente concluyentes (pruebas de hipótesis probabilísticas), además de la disposición de instrumentos que nos permiten describir, comparar, medir correlaciones y hacer predicciones.

De esta manera la Estadística permite, mediante la aplicación de un conjunto de reglas de razonamiento, pasar de proposiciones particulares a proposiciones generales, esto no significa, de manera alguna, que permita pasar de conceptos acerca de hechos sociales a leyes acerca de la esencia de los procesos de los cuales esos hechos o fenómenos forman parte. Esto último es tarea de la inferencia científica y no de la inferencia estadística.

En síntesis, podemos decir que es una forma de pensar pero no la "única manera científica de pensar" y más aún, encierra el peligro de ser aplicados inmoderadamente a cualquier contenido insustancial que, revestido ahora de la "formalidad y complejidad matemática" sirve para la extracción de conclusiones abusivas y absolutas- que nada aportan al conocimiento de los procesos sociales en cuanto a su génesis, su esencia, su estructura y a su desarrollo.

Ejemplo de lo anterior, se puede apreciar en ciertas prácticas investigativas, en donde aspectos del todo sociales se transforman en categorías particulares e independientes (*Factores*). Los diversos aspectos del todo sociales llegan de ese modo a aislarse y a transformarse en abstracciones formales indiscutibles y absolutas. Hagamos un breve paréntesis y hablemos un poco de la necesaria distinción que hay que hacer en el análisis e



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

interpretación de los hechos sociales: la distinción entre el manejo de los "factores" y el manejo de la concepción de "estructura.

Veamos en la práctica como se expresa esta situación. La distinción entre estructura económica y factor económico da lugar a la comprensión del significado central de la economía política en el sistema de las ciencias sociales y a la del primado de la economía en la vida social. Y es en base en el elemento económico y en la forma en que se lo conceptualiza y utiliza metodológicamente, como podemos establecer una primera y muy básica distinción que ejemplificaremos con dos investigaciones en el tema de la mortalidad infantil.

En el trabajo "Mortalidad en la niñez" de R. Puffer y Serrano, se señalan entre otros objetivos "estudiar las interrelaciones de las enfermedades y otros factores causantes de la mortalidad". Obviamente, entre otros esos "otros factores" aparece el socioeconómico, que en este trabajo es estudiado fundamentalmente a partir del "indicador nivel educacional de la madre". Por eso no es extraño que la principal conclusión sea: "...las madres sin ninguna instrucción se encuentran con más frecuencia en las reas rurales y una elevada proporción de sus hijos sufre de deficiencia nutricional"; Por consiguiente, de acuerdo a los resultados-recomendaciones de esta investigación, "elevar los niveles de educación de..." podría ser una de las soluciones al problema planteado.

Este tipo de conclusiones, en el cual los conceptos son numerizados y en la cual el contenido solo expresa una descripción rigurosa de los hallazgos derivados de la cuantificación de las variables investigadas (lectura de una tabla de distribución de frecuencia), es característico de una "disciplina" concebida solo como instrumento necesario para describir la estructura y el curso de los procesos sociales o como una especie de percepción cuantificada de la sociedad que se nos presenta numerizada. En este tipo de investigación se termina en la sustitución del análisis de la realidad por el de los signos, simplificando de esta manera, no solo la metodología de la investigación, sino la realidad misma;

Ejemplo de lo anterior, puede ser la siguiente Tabla simplificada sobre mortalidad infantil:

TABLA 1. Mortalidad en menores de un año. Valle del Cauca - 1990

Causas	Nº de Casos	%	Tasas X 10.5
Infeción intestinal	560	22.5	798
Trastornos relacionados Duración corta gestación y otras asociadas	276	11.1	393
Otras deficiencias nutricionales	256	10.3	365
Bronconeumonía y Asoc.	208	8.4	296
Hipoxinia y asociadas	180	7.2	256
Septicemia	155	6.2	221
Otras afecciones del Periodo perinatal	76	3.1	108
Síndrome Respiratorio	66	2.7	94
Meningitis bacteriana	65	2.6	93
Otras afecciones respiratorias del feto y recién nacido	52	2.1	74
Resto	591	23.8	842
TOTAL	2.485	100.0	3.542

En este tipo de Tablas, se cuantifican, según su impacto, una serie de "factores" o causas de mortalidad de menores de un año, sin especificar ni "jerarquizar" las causas básicas que determinaron los factores de morbilidad.

Otro trabajo sobre el mismo tema: ,l de Hugo Behm, "Mortalidad Infantil y nivel de vida", si bien inicia su análisis a partir del enunciado de un grupo de "factores económico-sociales", sobrepasa esta concepción y compara la mortalidad infantil entre grupos de "obreros" y "no-obreros.", Logrando así penetrar en la base de la estructura social, al señalar como variable independiente al modo de inserción de la población en las relaciones de producción.

En este trabajo, las conclusiones se presentan en forma diferente:

- "las relaciones de la mortalidad infantil con la clase social se estudian clasificando los nacimientos y defunciones en menores de un año de edad nacidos de acuerdo con la ocupación del padre del nacido vivo. Se establecen dos grupos: "obreros" y "no-obreros";
- La mortalidad infantil en la clase obrera es muy alta (126 por mil) y mayor que en el grupo "no-obrero" (67 por mil) [...]

En síntesis, en este punto nos interesa destacar que uno de los riesgos que debemos evitar en la investigación social es el de construir largas listas de factores o variables, sin una jerarquización explicativa; la realidad social jerarquiza las causas y el investigador debe evitar caer en la construcción de modelos matemáticos equilibrados, de una realidad que, insistimos, es en esencia desigual y desequilibrada. Esta diferenciación entre concepciones positivistas que cifran su base explicativa en la teoría de los factores y las concepciones que basan su explicación en una "estructura económica" que subordina el resto de los factores es, hay que reconocerlo, un corte que aún divide a los investigadores.

Y, ya que estamos hablando de diferentes "formas de leer", veamos con un ejemplo concreto, los milagros que se pueden hacer con una lectura "científica" de la información;

en el periódico El Tiempo del 24 de Mayo de 1992, se encuentra un artículo titulado "Lo que dicen las encuestas para un analista de opinión" de Carlos Lemoine:

"La encuesta del Centro Nacional de Consultoría mostraba que el 42.8 por ciento tenía una opinión positiva de la gestión de Gaviria. En enero de 1982, a los dos años de iniciado el mandato de Ronald Reagan, solo el 38 por ciento de la población consideraba positiva su gestión.

A la gestión Tachter, en marzo de 1982, un año después de comenzado su gobierno, la apoyaba el 34 por ciento de la población. Para el presidente Mitterand, en septiembre de 1985, el nivel de aceptación era el 41 por ciento y el de rechazo 48 por ciento. Esto demuestra que Gaviria tiene un nivel de apoyo un poco superior al que tuvieron Ronald Reagan, Margaret Tachter y Francois Mitterand cuando llevaban recorrido un lapso similar de tiempo en el poder.

Los "milagros" en la lectura van desde la comparabilidad no solo desfasados en el tiempo y el espacio hasta los malabares de intentar pretender como parámetro de comparación realidades socio-políticas totalmente ajenas. Esto es el reflejo de lo que la periodista Maria Teresa Herrán llamó en una reciente conferencia en el CIDSE, "el gobiernismo a ultranza de los analistas del acontecer nacional." Finalmente, en esta breve introducción al análisis de datos cuantitativos, debemos destacar que el juicio crítico no lo estamos formulando solo acerca de la corrección o no del manejo estadístico sino que pretendemos insistir en que lo importante reside en la vinculación que tenga la investigación con una concepción o perspectiva teórica. Y tal como se mencionó previamente, la corrección de una concepción teórica reside en la inclusión del componente histórico, del económico, del político, del demográfico, de lo cultural, de lo ideológico, etc. etc., que articulan los desiguales procesos sociales según clases sociales.

ANÁLISIS E INTERPRETACION DE DATOS.

El término *análisis* aquí lo vamos a entender, primero como un proceso de búsqueda de elementos o aspectos o configuraciones que caractericen los procesos o hechos estudiados. Y segundo, una "lectura" de datos, de acuerdo a criterios u orientaciones de clasificación y ordenamiento.

Cuando se trabaja con un cúmulo de información (datos) asociada a una población, surgen varias interrogantes con relación al "comportamiento de los datos".

Sobre el "comportamiento de los datos" de una población, la tarea de un analista puede enfocarse en tres líneas:

1. *La descripción del dato* (análisis descriptivo): que consiste en una serie de técnicas que permiten presentar en forma resumida la información obtenida;

2. *la Inferencia estadística*: tiene como finalidad examinar la posibilidad de "generalización" de los resultados a una población. Comúnmente se encuentran 2 tipos de Análisis Inferencial:

2.1. La Estimación: con base en resultados muestrales (estadísticos) se logran conclusiones aplicables a la población.

2.2. Las Pruebas de Hipótesis: cuyo objetivo es aportar información que permita aceptar o rechazar alguna suposición o respuesta tentativa acerca del "comportamiento de las variables" en la población estudiada.

3. *La asociación de variables*.

OBJETIVO DE LA UNIDAD.

La presentación de las diversas técnicas de carácter estadístico destinados a "leer" datos cuantitativos, pone énfasis en la lógica y en la oportunidad de su uso: ¿Qué estadísticos usar y cual es el significado de los estadísticos resultantes?

Nota metodológica. La dirección del análisis o la búsqueda-lectura se hace dentro de ciertas "orientaciones" que son señaladas generalmente por los objetivos del estudio.

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

El análisis de datos, en el campo de los diseños cuantitativos, en términos generales según Briones_ se dirige en este campo a encontrar:

- 1) elementos, estructuras o configuraciones que caractericen, en términos cuantitativos, a los colectivos o procesos en los cuales se han definido problemas de investigación;*
- 2) diferencias entre ellos;*
- 3) asociaciones y correlaciones simples y complejas;*
- 4) clasificaciones y topologías y*
- 5) Asociaciones y correlaciones relacionadas con hipótesis causales."*

Nota metodológica: en la lectura sociológica de una información, es válido reconocer el papel que juega el análisis estadístico, como instrumento auxiliar del análisis de datos.

En el análisis hay tres conceptos básicos que sirven para "leer" una información cuantificada:

1. UBICACION O LOCALIZACION_ DE LA INFORMACION

Descrita por medidas de tendencia central tales como la Media, Mediana y la Moda. Para lo cual se calculan:

- 1.1. Valores máximos -mínimos de los datos.
- 1.2. Valores promedio
- 1.3. Frecuencias o NS de individuos (U. A.)
- 1.4. Proporciones - porcentajes
- 1.5. Comparación gráfica de varios grupos

2. LA DISPERSION DE LOS DATOS RESPECTO A SUS PUNTOS DE AGRUPACION Y LOCALIZACION.

Para su estudio se calculan coeficientes, como:

- 2.1. Varianza
- 2.2. Desviación estándar
- 2.3. Rango de variación

3. LA ASOCIACION DE COMPORTAMIENTOS DE VARIABLES (COVARIACION)

La asociación entre variables es descrita por los:



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

3.1. Coeficientes de correlación.

3.2. Modelos de regresión.

1. ANALISIS DESCRIPTIVO DE LOS DATOS

La primera tarea del análisis consiste en obtener una información sobre las características o propiedades (variables del estudio) de las unidades de análisis.

Una segunda tarea hace referencia a "la comparación de sub - grupos en relación con una o más características que son importantes en relación con una o más características que son importantes en los objetivos formulados por el investigador."

USOS DEL ANALISIS DESCRIPTIVO:

1. El comportamiento de cada variable, es decir, cómo se presentan o varían los datos;
2. El punto alrededor del cual se tienden a agrupar los datos y,
3. La magnitud de la dispersión;

Nota metodológica. El tipo de análisis descriptivo que debe emplearse en cada variable depende del nivel en que fueron tomados o medidos los datos._

I. ANALISIS DE LA DISTRIBUCION DE LAS FRECUENCIAS DE UN CONJUNTO DE DATOS.

Al analizar el comportamiento de una variable, básicamente el investigador esta interesado en:

- Hacer una lectura global de los datos (leer la distribución de los datos: frecuencia absoluta o relativa; medidas de tendencia central, etc.)

-Buscar una explicación de la variabilidad de la característica, propiedad o variable.

Fundamentalmente hay dos preguntas que deben responderse:

2.1. Cómo varían las categorías dentro de la variable (análisis marginal de los datos).

2.2. Qué, factores están asociados a esta variabilidad (cruce de variables).

Ejemplificación:

Grados de aceptación familiar del embarazo

Desacuerdo-rechazo	80.96%
Acuerdo o aceptación solidaria	11.90%
Indiferencia o aceptación obligada	7.14%

Tipo de "relación" que tenían entre las jóvenes con el padre de su hijo.

Novio	85.68%
Amigo	7.14%
Conocido ocasional	4.76%
otro	2.38%

Nuevo estado de su relación afectiva

Relación terminada por

Abandono	59.50%
Permanencia del compañero	14.28%

Relación terminada por

Consenso	11.90%
----------	--------

No responde	9.52%
Otros	4.76%

Análisis de datos. Al analizar los gráficos y cuadros que hacen referencia a las vivencias de estas jóvenes que aquí se denominan "embarazadas precoces", la primera conclusión a la que se llega es que estos adolescentes han asumido una vida sexual sin saber como hacerlo racionalmente -el 80 por ciento de ellas no utilizó ningún método anticonceptivo-, prácticas que son rodeadas de tabúes y mitos que los lleva a tomar muchos riesgos, a sufrir la hostilidad y desconfianza de su medio familiar e incluso afectivo -esto se aprecia en el altísimo grado de rechazo familiar (85%) y el abandono de sus compañeros o novios y término de la relación original: Alrededor de 2 de cada 3 abandonan a su compañera embarazada o lactante. (71.4%). No olvidemos el tipo de relación que tenían las jóvenes con el padre de su hijo. En un 85 por ciento, éste había sido el novio, es decir, había sido una relación con un fuerte contenido emocional y afectivo. _

II. ESTADÍSTICAS PARA MEDIR LA LOCALIZACIÓN Y LA VARIABILIDAD DE UN CONJUNTO DE DATOS.

2.1. Estadísticas de localización.

2.2.1. Para medir la localización de los datos en una "escala nominal" _ se utiliza el valor de la variable que tiene mayor frecuencia o agrupación, este valor se conoce como la Moda.

Ejemplificación:

Supongamos que estamos interesados en estudiar conocimientos y actitudes en la rea de la sexualidad en un grupo de niñas de un Colegio. Obviamente, es importante para los efectos del análisis del hecho, la edad de los sujetos del estudio.

¿La pregunta básica que uno se hace en el intento de describir globalmente el Grupo es "cual es la edad más representativa en el grupo?" Hay tres (3) respuestas a esta pregunta que corresponden a 3 estadísticos que describen la "localización" de los estadísticos de tendencia central. A simple vista se puede afirmar que la edad representativa es 7 años, ya que la mayoría declaró esa edad; podríamos afirmar que la Moda es igual a 7 ya que es el valor de la distribución que aparece con mayor frecuencia.

Veamos que nos muestra la Tabla 1:

Tabla 1 Distribución según grupos de edades.

Grupo A. estudiantes mujeres Colegio Alemán

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

Para medir el punto que está al centro de la distribución se utiliza la Mediana, es decir el valor que es mayor o igual que la mitad de los datos y menor o igual que la otra mitad. O, la medida que divide a cualquier distribución en 2 partes iguales.

En el caso particular del Grupo A, la mediana es 7.17

Nota metodológica: para calcular la mediana se tienen que ordenar los datos y determinar ,la que está a la mitad. Sirve para datos tomados de escalas de medición Nominales, Ordinales e Intervalos

2.2.3. En el lenguaje cotidiano, siempre se habla de promedios (promedios aritméticos); de acuerdo a las edades declaradas en la Tabla 1. Podemos decir que el promedio es de 7.25 años. Este promedio es el que se llama Media Aritmética y se calcula así:

$$\text{Círculo de la Media (X)} = \frac{\text{-----}}{N} = \frac{\text{-----}}{8} = 7.25$$

NOTA METODOLÓGICA: La Media es una medida de localización que fija el promedio de los datos. Cuando los datos han sido tomados en una escala de Intervalos, además de la Moda y la Mediana se puede calcular la Media aritmética:

III. ESTADÍSTICAS PARA ANALIZAR LA DISPERSION DE UN CONJUNTO DE DATOS.

Con las medidas de localización se puede caracterizar un conjunto de datos, esta caracterización depende de la variabilidad de los datos alrededor del punto de localización. Cuando los datos se toman en una escala de Intervalos hay 3 medidas de dispersión de uso frecuente en la práctica estadística:

1. RANGO: que se define como la "diferencia o distancia" entre el dato mayor y el menor.

2. **VARIANZA:** que se define como el promedio de los cuadrados de las diferencias de cada dato con respecto a la "Media" _.

3. **DESVIACION ESTANDAR:** Se define como la dispersión con respecto a la Media.

Ejemplificación: Supongamos que vamos a comparar la distribución de edad de dos grupos de estudios sobre sexualidad de un colegio.

Tabla 2 Distribución según grupos de edades de 2 grupos de estudiantes sexo femenino Colegio Alemán.

Edad (X)	Grupo A	Grupo B
	Frecuencia	Frecuencia
11	0	1
10	0	0
9	1	1
8	2	1
7	3	3
6	2	0
5	0	1
4	0	1
Total	8	8
MODA	7	7
Media	7.25	7.25



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

Análisis: una primera lectura nos muestra que el Grupo A es más homogéneo, en términos de la distribución de las edades, en cambio, el Grupo B, podría ser considerado como heterogéneo (mayor dispersión de edades. En las Ciencias Sociales, el estadístico más utilizado para medir la dispersión es la Desviación Estándar.

Variación y dispersión de las edades del Grupo A.

Edad (X)	Frecuencia (f)	fX	fx ²
9	1	9	81
8	2	16	128
7	3	21	147
6	2	12	72
Total	8	58	428

Cálculo de la Varianza y Desviación estándar

$$(fX)^2 = (58)^2 = 3364$$

$$\frac{(fX)^2}{N} = \frac{3364}{8} = 420.5 \quad fX^2 = 428$$

$$fX^2 - \frac{(fx^2)}{N} = 428 - \frac{450.5}{8} = 7.5$$

$$S^2 = \frac{1}{N} \left[fX^2 - \frac{(fX)^2}{N} \right] = \frac{1}{8} (7.5) = .94$$

$$S = S^2 = .94 = .97 \text{ años}^3$$

La lectura del estadístico muestra que el Grupo A es muy homogéneo en términos de edad.

IV. COMPARACION GRAFICA DE VARIOS GRUPOS.

Si dentro de una investigación se quieren comparar varios grupos en términos de una variable, el Histograma de Frecuencias y los gráficos esquemáticos (BOX PLOTS) son una herramienta útil.

Los gráficos esquemáticos permiten visualizar varios aspectos de una Distribución de Frecuencias:

- la localización (Mediana)
- La variabilidad (distancia intercuartil)
- Simetría.

VI. COMPARACION DE COLECTIVOS O DE CONJUNTO DE UNIDADES DE ANALISIS.

En el análisis de datos, la comparación de grupos, sub-grupos, un mismo grupo en diferentes dimensiones espacio-temporal, hechos o procesos juega un papel importantísimo.

5.1. COMPARACION DE PORCENTAJES.

La técnica consiste en comparar los porcentajes tabulados sobre cada uno de los totales que constituyen los sub-grupos, con el objetivo, en la mayoría de los casos, de señalar diferencias entre ellas. Donde mejor se aprecia el concepto es en la lectura de "un cruce de variables y comparación de variables en una tabla o cuadro multivariados."

Nota metodológica:

- a) La lectura en una tabla multivariada depende en que, dirección se han calculado los porcentajes;
- b). Los porcentajes deben calcularse en una dirección de la variable que consideramos como causal.

Ejemplificación:

Tabla 5 Factores de deterioro social y ambiental en Medellín por estratos socioeconómicos.

En porcentaje %

Factores Estratos	BAJO	BAJO MEDIO			MEDIO	MEDIO ALTO	TOTAL
	BAJO	BAJO	MEDIO	ALTO			
Inseguridad	11.1	33.6	22.2	40.9	59.0	80.0	30.8
Polución y Cong.	16.7	15.5	15.3	13.2	7.7	0.0	14.5
Ambiente social	33.3	11.2	12.1	10.3	5.1	0.0	11.8
Modo ser gente	2.8	7.6	10.5	5.2	2.6	0.0	8.1
Lugar determin.	2.8	2.8	8.6	7.9	0.0	10.0	7.7
Serv. Publ. Soc.	11.1	4.7	8.6	7.4	0.0	0.0	7.0
Todo	2.8	1.4	1.3	1.7	0.0	0.0	1.4
Nada	19.4	20.2	21.4	13.4	25.6	10.0	18.7
Totales	100	100	100	100	100	100	100

Análisis de datos: (de una columna) en el estrato medio, una de cada 5 personas, declaran que la inseguridad es un factor principal del deterioro social y ambiental en la ciudad de Medellín; no así, en el estrato bajo-bajo, en donde este factor pasa a un plano secundario tomando fuerza la apreciación que es el ambiente social (33.3%) ,l que influye en el proceso de deterioro.

5.2. COMPARACION CON MEDIDAS DE CONCENTRACION.

Esta técnica permite "comparar" el grado de concentración (o, desigualdad) que se puede dar dentro de una población o grupo social respecto de una característica o propiedad determinada, como podría ser: nivel educativo, nivel de ingresos, nivel de consumo, nivel de participación ciudadana, etc.

Nota metodológica: Índice de Concentración o Coeficiente de Gini; es una medida del grado de concentración. Sus valores van de cero (0) a uno (1);

Significado de los valores:

CERO (0) si se da una igualdad máxima o no existe concentración;

UNO (1) Si solo existe una "completa concentración" o "absoluta desigualdad".

Ejemplificación: Se tiene la nómina e los docentes de la UniValle, ordenada en orden descendente y se quiere determinar la distribución de ingresos en el sector docente.

niveles ingresos	personas en cada estrato	personas %	total ingresos	ingresos %
14 - 13 **	30	3.4	785	7.4
12 - 11	92	10.5	2.450	23.1
10 - 9	105	11.9	3.120	29.5
8 - 7	275	31.2	3.320	31.4
6 - 5	253	28.7	505	4.8
4 - 0	125	14.3	410	3.8
TOTAL	880	100.0	10.590	100.0

** Salario Mínimo 1992 = \$ 65.400.

niveles	personas	ingresos	$X_{i-1} * Y_i$	$X_i * Y_{i-1}$
(X)	(Y)			

14 - 13	3.4	7.4	103.7	102.9
12 - 11	13.9	30.5	834.0	960.8
10 - 9	25.8	60.0	2.879,1	3.426,0
8 - 7	57.1	91.4	5.493,0	7.842,1
6 - 5	85.8	96.2	8.580,0	9. 620,0
4 - 0	100.0	100.0		

17.889,8 21.951,8

Utilizando estos valores en la fórmula del Coeficiente de Gini se logra lo siguiente:

Coeficiente de Gini= $\frac{17.889,8 - 21.951,8}{10.000}$

= 0.41

El valor calculado de 0.41 indica una concentración mediana en el sector docente de la UniValle, es decir, no existiría gran desigualdad en la distribución de los ingresos entre los docentes.

Nota metodológica: Un $G = 0.50$ puede indicar una concentración media de los ingresos de un sector, es decir, no existiría una gran desigualdad entre los sujetos (unidades de análisis) del estudio.

5.4. COMPARACION DE PROMEDIOS.

Nota metodológica: 1. Supongamos una Población desagregada por sub-grupos.

2. Comparación de las variaciones de porcentajes en los dos sectores poblacionales.

Área Metropolitana CALI: niveles promedio de ingresos de la PEA nativa y migrante por quintiles.

Quintil	Nativa	Inmigrante	Variación
---------	--------	------------	-----------

I	4.0%	3.1%	+ .9
II	10.3%	8.8%	+ 1.5
III	14.6%	16.0%	- 1.4
IV	20.8%	22.4%	- 1.6
V	50.3%	49.7%	+ .5

5.3. COMPARACION DE TASAS.

Nota metodológica: frecuentemente dentro del análisis descriptivo se comparan tasas o índices de diferente naturaleza.

Ejemplificación: "Comparación de tasas de deserción escolar, calculadas con diagrama de flujo de una cohorte que inició estudios de educación básica de 8 años.

Tabla 25 Comparación de tasas de deserción escolar

	Total nacional	Total rural
	%	%
Nacional y Municipal	33.2%	39.2
Particular subvencionada	14.8	24.6
Part. no-subvencionada	9.9	10.2

“La mayor deserción entre los subsistemas de la educación básica la tiene la educación nacional y municipal de las reas rurales, con 39.2%, un poco superior al promedio nacional de 37.8%. En el conjunto urbano y rural, la cifra correspondiente es de 33.2%... además, estos valores quedan muy por encima de la educación subvencionada, con 14.8% a nivel nacional y 24.6% en las reas rurales. Las tasas más bajas corresponden, en ambos casos, a la educación no subvencionada". _

5.5. COMPARACION DE INDICES.

Supongamos que interesa analizar las tasas de movilidad ocupacional _ (cuantos años han trabajado desde su primera ocupación y cuántas ocupaciones han desempeñado desde entonces.

Ejemplificación:

Tabla. TASAS DE MOVILIDAD OCUPACIONAL ENTRE PUESTOS, SEGUN SEXO. _

Movilidad *	hombres	mujeres	total
Alta	19.0	33.9	23.9
Media	41.5	35.7	40.0
Baja	39.5	30.4	37.1
TOTAL	100.0	100.0	100.0
	(774)	(277)	(1.051)

* Las siguientes son las definiciones de las tasas de movilidad: Alta movilidad: más de 5 cambios por d,cada; Media movilidad: 2 a 4 cambios por d,cada; Baja movilidad: 1 a 1.9 cambios por d,cada.

La información de la Tabla anterior muestra que del total de las mujeres una de cada tres tiene un historial en el rango de la Alta movilidad ocupacional; en cambio en los hombres, la situación parece ser la contraria: sólo un 19 por ciento mostró una alta movilidad ocupacional y casi el 40 por ciento declaró una baja movilidad ocupacional.

5.6. COMPARACION EN MUESTRAS.

Problema: Intentar comparar los valores encontrados (diferencias) en sub-grupos producto de una población muestral.

Solución: Efectuar "pruebas de significancia estadística para pronunciarse sobre la "significación estadística de las diferencias encontradas".

Nota metodológica: Prueba especial: Test de Significación t. de Students.

La t. de Student permite verificar la significación estadística de la diferencia que puede darse entre 2 medios aritméticos correspondientes a otros tantos grupos.

La significación estadística no reemplaza la significación científica en términos de la evaluación del conocimiento o de la significación práctica. Por ejemplo, "las diferencias entre grupos pueden ser estadísticamente significativas pero científicamente triviales y sin valor práctico.

Ejemplificación: Se tienen dos muestras poblacionales de municipios, una de municipios urbanos y otra de municipios rurales y se quiere investigar si existen diferencias significativas en las preferencias electorales de los habitantes de esos tipos de municipios. En las dos muestras se calcula el porcentaje de voto conservador y se obtiene la siguiente distribución:

Municipios urbanos: $N_1 = 23$ $X_1 = 43 \%$ $S_1 = 12 \%$

Municipios rurales: $N_2 = 41$ $X_2 = 51 \%$ $S_2 = 15 \%$

El planteamiento estadístico del problema es:

$$H_1 = \mu_1 \neq \mu_2 \quad X_1 \neq X_2$$

$$H_0 = \mu_1 = \mu_2 \quad X_1 = X_2$$

Con un nivel de significancia al .01

Utilizando la fórmula de t de Student:

$$t = \frac{43 - 51}{\sqrt{\frac{12^2 + 15^2}{43 + 51}}} = \frac{-8}{\sqrt{\frac{334 + 225}{98}}} = \frac{-8}{\sqrt{\frac{559}{98}}} = \frac{-8}{2.78} = -2.87$$

Grados de libertad = 43 + 51 - 2 = 92

En la Tabla de Distribución t se aprecia que para rechazar H_0 necesitamos obtener una t igual o superior a 2.37. Como en efecto se obtiene un valor superior (2.87), se rechaza la Hipótesis Nula H_0 y se confirma que si hay diferencias en el comportamiento electoral según el tipo de municipio.

V COMPARACIONES MÚLTIPLES.

En los diseños cuantitativos, es posible en ciertas ocasiones, que se den situaciones en las cuales es necesario "comparar 3 o más medios aritméticos que corresponden a diversos valores de una misma variable dependiente afectada por una o más variables independientes."

5. 1. ANALISIS SIMPLE DE LA VARIANZA.

Problema: analizar la validez (significación estadística) de unos resultados producto de tres factores causales.

Cuando se requiere verificar la significación estadística de la diferencia que puede darse entre dos MEDIOS aritméticos (mediante la Prueba t), correspondientes a otros tantos grupos, Briones afirma lo siguiente:

...en el primer caso, el ejemplo típico sería comprobar la significación estadística de los rendimientos en el aprendizaje de una cierta materia enseñada por tres m, todos distintos; en el caso de más de una VARIABLE INDEPENDIENTE, analizar los mismos resultados considerando la acción conjunta de la variable m, todo de enseñanza y otra que podría ser la de "ambiente escolar", con dos o tres categorías diferentes. Cuando se desea resolver el problema de la comparación de tres o más medios aritméticos de una VARIABLE DEPENDIENTE sobre la cual incide una o más variables independientes, la técnica estadística utilizada es el ANALISIS DE VARIANZA." _

Ejemplificación: "Se desea saber si tres métodos de enseñanza producen resultados diferentes y estadísticamente significativos en el aprendizaje de ciencias sociales, medido con una prueba en base a puntaje. Se conforman tres grupos de estudiantes seleccionados independientemente, al azar, de una escuela, con los puntajes obtenidos después que cada uno de ellos fue enseñado por el método que se indica.

A Método de lectura individual	B. Método expositiva	C. Método de discusión
Puntajes		
6	6	2
9	7	5
11	9	4
10	8	6
X = 9.0	X = 7.5	X = 4.25

PREGUNTA DEL PROBLEMA: ¿Son las diferencias observadas en los 3 medios aritméticos estadísticamente significativos?

Nota metodológica: El análisis de la varianza simple que se utiliza para hacer este tipo de comparaciones se siguen los siguientes pasos:

- formulación de una hipótesis nula;
- formulación de una hipótesis alternativa;
- elección de un nivel de significación;
- elección de una distribución de muestreo (Distribución F);
- cálculo de la estadística F;
- decisión basada en el valor encontrado en la Tabla F.

2. EL ANALISIS INFERENCIAL

I. LA ESTIMACION.

Algunos trabajos estadísticos tienen como objetivo la "estimación" de uno o varios "parámetros" de una población. Con este objetivo se toma una muestra y con base en los "resultados muestrales" (parámetros estadísticos) se logran conclusiones aplicables a la población.

La base de un procedimiento de estimación radica en el hecho de que para estimar un parámetro poblacional o estimativo se usa una muestra.

Nota metodológica: El proceso de generalización a una población es conocido como la INFERENCIA ESTADISTICA.

LA DISTRIBUCION NORMAL: Anteriormente se presentaron algunos procedimientos del análisis descriptivo que nos permita "ver" el comportamiento de la distribución de unos datos. Veamos ahora para que sirve la llamada "Distribución Normal": _

- Por sus características, es un instrumento esencial en la inferencia estadística y la prueba de hipótesis;
- Su ventaja es que nos permite a comparar distintas distribuciones de frecuencias.

Nota metodológica: para la representación gráfica de la curva normal con una media y desviación estándar determinada, en el programa STATGRAPHICS utilizar DISTRIBUTIONS FUNCTIONS + DISTRIBUTION PLOTTING.

Ejemplificación: "En la Tabla 3.1., Se presentan los porcentajes de individuos con resultados inferiores a un puntaje dado en la prueba de aritmética y los porcentajes que se obtendrían al suponer una distribución normal.

Tabla 3.1 Porcentaje de individuos con puntaje menor que el indicado en la prueba de aritmética

Puntaje	% en la curva	% real en la Prueba
	Normal	de Aritmética
2.54	2.3	1.3
4.40	15.9	15.6
5.00	21.0	29.4
6.00	37.4	47.6
8.72	84.1	77.9
9.00	87.2	90.9
10.78	97.7	98.

"El lector puede verificar que, en la curva normal, entre los puntajes que distan una desviación estándar de la media (4.40 y 8.72) esté el 68.2% de los individuos y entre los que se separan dos desviaciones estándar (2.54 y 10.78) el 95.4 por ciento.

En una tabla normal se puede verificar que entre los que distan 3 desviaciones estándar de la media estén el 99.7 por ciento. En una variable con distribución normal prácticamente todos los valores observados deben estar a menos de "tres desviaciones" estándar de la Media. En otras palabras, si aparece un dato mayor que la media en más de tres desviaciones es muy probable que la información de este individuo tenga algún error o este pertenezca a una población diferente a la del resto de los individuos, por ejemplo, a una población con un puntaje promedio mayor."

ESTIMACION DE LA MEDIA DE UNA POBLACION:

La media de una variable de una población se puede estimar por la media de una muestra (media muestral) calculada de dicha población.

Ejemplificación: En un estudio muestral sobre distribución de ingresos entre los residentes de un barrio clasificado en el estrato Medio Medio -3-; Se encontró que el ingreso familiar medio de los residentes fue de \$ 150.000.

Aquí surge un problema metodológico, en la posibilidad de generalizar este resultado a todos los habitantes del barrio, acerca de la certidumbre si el ingreso medio calculado (la media muestral obtenida) diferir del ingreso medio del total de la población residente. La respuesta la encontramos, de alguna forma, en una convención que determina el grado de confianza de que el cálculo obtenido de la muestra ser similar al de la población total. Generalmente, en las Ciencias Sociales, se habla de un grado de confianza de un 95%. En nuestra ejemplificación, tenemos una muestra de 1.000 residentes en el Barrio. El ingreso medio fue calculado en \$ 150.000 y la desviación estándar en \$20.000.

Calculemos el error estándar mediante la siguiente fórmula:

$$S_x = \frac{20.000}{\sqrt{1000}} = 632.3$$

Con este resultado, se pueden hacer algunas estimaciones, por ejemplo, si se está trabajando al 95% de nivel de confianza (1.96), el rango de los valores estar alrededor de: \$ 150.000 ± 1.96 (632.3) = \$ 150.000 ± \$ 1.239.3.

A partir de estos cálculos, se puede inferir que el ingreso medio en el Barrio varía entre \$ 148, 760,7 y \$ 151.239.3

II. PRUEBA DE HIPOTESIS

Cuando se está trabajando con diseños cuantitativos, las variables pueden tomar características cuantificables, permitiendo así la aplicación de técnicas estadísticas, que nos van a permitir analizar el comportamiento de los datos reunidos (localización, dispersión y asociación de los datos). Una de estas técnicas es la Prueba de Hipótesis" que permite probar o disprobar una suposición o afirmación hipotética (Hipótesis de investigación o de trabajo).

ETAPAS DEL PROCESO DE PRUEBA DE HIPOTESIS:

1. Planteamiento de la hipótesis de trabajo o investigación. Al mismo tiempo hay que definir una hipótesis nula (H_0) que debe plantear que no existe la relación entre las dos características o propiedades.

Existen dos tipos de hipótesis estadísticas:

- Hipótesis Nula (H_0)

- La Hipótesis de investigación, de trabajo o Alternativa (H_1)

1. Obtener la distribución de muestreo: Es la distribución de una muestra de medios a partir de muestras.
2. Seleccionar un nivel de significancia y una región de rechazo.
3. Realizar la prueba estadística.
4. Comparar -aceptar o rechazar la hipótesis de trabajo- y sacar conclusiones.

LA ESTADISTICA NO PARAMETRICA

Aquí se ha afirmado que los niveles de medición de las variables o características que se miden condicionan el tipo de análisis estadístico que se debe utilizar en la interpretación. Hay datos que en términos generales reúnen 3 características: nivel de medición nominal u ordinal, pequeños tamaños muestrales y una distribución de la población diferente a lo normal. Cuando la distribución no es normal se denomina NO-PARAMETRICA y se utilizan técnicas que se denominan técnicas no-paramétricas. El resultado solo muestra si la relación es significativa o no y no se pretende proyectar los resultados a la población.

PRUEBAS ESTADISTICAS PARA VARIABLES NOMINALES

Generalmente en el proceso de ordenamiento y clasificación de la información, muchas de las características (variables) que estudiamos no pueden ir más allá de la simple clasificación. Las pruebas estadísticas con medidas a este nivel nominal se denominan Pruebas de Contingencia. Una de estas técnicas de amplio uso en el análisis de datos es la Prueba de Chi-cuadrado (χ^2).

Es un estadístico que permite analizar si unas frecuencias obtenidas difieren o no de las que se esperaban bajo ciertos supuestos teóricos.

Ejemplificación. Se quiere averiguar si existe una relación entre los niveles de integración social de grupos familiares que residen en un particular tipo de viviendas, casa unifamiliar tradicional o tipo apartamento localizados en unidades residenciales cerradas.

Hipótesis de trabajo: El tipo de vivienda en residentes urbanos de ciudades intermedias influencia ciertos comportamientos sociales de integración.

La hipótesis nula es que no existe diferencia entre el tipo de vivienda y el tener o no tener un comportamiento integrativo.

TABLA

FAMILIA	Tipo de vivienda	
	casa	apartamento
Alta integración	32 a	14 b
Baja integración	28 c	46 d
		N= 120

$$a = (a+c) (a+b) \div N = 23$$

$$b = (b+d) (a+b) \div N = 23$$

$$c = (a+c) (c+d) \div N = 37$$

$$d = (b+d) (c+d) \div N = 37$$

Familia	Tipo de Vivienda	
	Casa	Apartamento
Alta integración	23	23
Baja integración	37	37

$$X^2 = \frac{(32-23)^2}{23} + \frac{(14-23)^2}{37} + \frac{(28-37)^2}{37} + \frac{(46-37)^2}{37} =$$

$$X^2 = 3.52 + 3.52 + 2.19 + 2.19 = 11.42$$

El X^2 es igual a 11.42.

Lectura de los resultados: ¿Hay o no relación entre las dos variables?

1 Se determina el valor crítico en la Tabla de la Distribución de X^2 .

2§ Se calculan los grados de libertad: $(2-1) (2-1) = 1$;

3 Se supone un nivel de significancia de .05

4 De acuerdo a la Tabla, se necesita un nivel de 3.841 o mayor para rechazar la H_0 ;

5 el calculado $X^2 = 11.42$, es mayor a 3.84, entonces se rechaza la H_0 ;

6 Se aprueba que la Hipótesis de trabajo es válida y se concluye que si hay diferencias entre las dos variables.

3. METODOS PARA ANALIZAR RELACIONES CON DATOS EN ESCALAS CUANTITATIVAS.

Uno de los objetivos del análisis de una información es detectar si existen relaciones entre el objeto o característica principal de estudio y otros objetos o características. Briones _ afirma que "se acostumbra a 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."

Nota metodológica: La búsqueda de posibles relaciones entre dos variables debe iniciarse con un análisis gráfico. El resultado solo mostrar si la relación es significativa o no.

Técnica: Los diagramas de dispersión.

En estos gráficos se puede analizar:

- la forma de la relación _
- la fuerza de la relación

Además de lograr saber si existe o no relación entre las variables, se puede calcular la fuerza de la relación. Esta puede ser débil o fuerte o inclusive puede no existir. Generalmente, si la relación entre las variables se acerca a cero (0) indicaría debilidad y si se acerca a uno (1) sería indicativo de una alta correlación.

Unas de las más utilizadas medidas de asociación con coeficientes no-paramétrico es el Coeficiente de Contingencia C; hay otras como el Q de Yule; Lambda; la tau de Goodman y Kruskal; el coeficiente Gama; el D de Sommer; etc.; que se utilizan dependiendo de los niveles de medición de las variables asociadas.

3.1. ANALISIS DE CORRELACION.

Una forma de analizar la FUERZA de la relación entre dos variables la proporciona "El Coeficiente de Correlación lineal o Pearson, para variables medidas a nivel ordinal. Simplemente mide el grado de asociación entre dos variables.

Nota metodológica. Lectura de un valor de un Coeficiente de Correlación de Pearson:

El "r" puede tomar un valor entre -1 y 1; el valor -1 significa que la correlación es inversa y, el valor +1, significa que la r es directa. Que la "r" sea directa, es decir que el aumento de una variable esta asociado con el aumento de otra, en este caso el "r" es positivo.

Si el crecimiento de una variable esta asociado al "decrecimiento" de la otra, la relación es inversa.

Que exista o no, una correlación significativa entre variables puede deberse en otras causas a la "variabilidad asociada al procedimiento de muestreo.

Un valor de "r" = 0.514696 puede entenderse como una correlación relativamente alta, entre dos variables.

Ejemplificación:

Tabla 3 Grados de satisfacción Actividad Concejo Municipal CALI y niveles de información política de los informantes.

Encuestados	Grado	Nivel	Orden 1	Orden 2	D	D ²
	Satisfac.	Información				
1	30	50	1	1	0	0.00
2	28	40	2	3.5	1.5	2.25
3	27	49	3	2	1	1.00
4	25	34	4	8	4	16.00
5	24	40	5	3.5	1.5	2.25
6	23	38	6	5	1	1.00
7	20	37	7	6	1	1.00
8	19	30	8	9	1	1.00
9	18	35	9	7	2	4.00
10	17	29	10	10	0	0.00

N =10

D² = 28.50

Nota metodológica: No olvidar que estamos tratando de averiguar cual es la FUERZA entre las variables "grados de satisfacción con la actividad del Concejo Municipal de Cali y el nivel de información política de los informantes.

Se aprecia que se ordenan los informantes de mayor a menor grado de satisfacción y nivel de información, de tal forma que la persona que tenga más información o más satisfacción se identifica con el número 1, la segunda con el N° 2 y la menor con el N° 10. Una vez

realizado el ordenamiento, queremos saber que tan cerca está el ordenamiento de una variable del ordenamiento de la otra.

En las Columnas de Orden 1 y Orden 2 se ordenan los puntajes de cada miembro de la muestra de informantes; si los ordenamientos coincidieron se lograría una relación perfecta y no sería necesario hacer el cálculo. En este ejercicio este no es el caso.

Desarrollo del ejercicio:

Se calcula la diferencia, en Columna D, entre cada par de ordenamientos.

Cada diferencia se eleva al cuadrado: Columna D²

Cálculo de la Correlación de Spearman:

$$r_s = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

$$r_s = \frac{6(28.50)}{10(102 - 1)} = 1 - .173 = .83$$

$$r_s = .83$$

Análisis del dato obtenido: Este índice calculado indica una Correlación fuerte y permite concluir que existe una relación entre nivel de información política de los informantes y el grado de satisfacción con la actividad del Concejo Municipal de Cali.

En este ejercicio se sabe que la correlación $r_s = .83$ se acerca bastante a 1 y es fuerte pero no se sabe cuan "significativa" es la correlación encontrada. Es decir, se sabe que existe pero no se sabe que tan significativa es. Para poder concluir acerca del Nivel de Significancia, se tiene que comprar el estadístico obtenido con la Distribución de los Valores críticos de r_s para cuando las muestras tienen un tamaño entre 4 y 30 casos.

En el caso del Ejercicio, N es igual a 10 ($N=10$) y al usar la Tabla de Distribución de los Valores Críticos de r_s nos encontramos con lo siguiente:

- 1 Tenemos una correlación con un valor de .83
- 2 Un r_s esperado para que sea significativo al .01 tiene que ser igual a .746;
- 3 Por lo tanto se concluye que la relación encontrada (.83) es significativa, al menos al Nivel .01

3.2. ANALISIS DE REGRESION.

Con esta técnica no solo se obtiene información de la existencia de una asociación entre variables sino que además información acerca de la naturaleza de la relación; es así como se entiende el concepto de "dependencia de una variable a partir de otra", por ejemplo, "rendimiento académico a partir de "lugar de residencia"; mortalidad infantil a partir del oficio de los padres; etc.

En el modelo de regresión se juega con el concepto de Variable Dependiente (variable que es la consecuencia de...), y de Variable Independiente que puede ser la causa o factor causal, es aquí donde la perspectiva teórica juega un papel importante ya que permitir definir cual es cual. _

En el modelo de regresión, la idea es que si existe relación entre variables, el conocimiento de una de ellas, nos puede ayudar a "predecir" el comportamiento de la otra y ,esta "predicción" se puede presentar así

$$Y = f(X)$$

Se trata de "explicar" los cambios de una Variable-Respuesta - Y - (Var. Dependiente) en términos o en función de una o varias variables explicativas - X - (Var. Independiente.

Nota metodológica: Si se tiene una sola variable Independiente, el modelo se llama "Regresión Simple"; si se consideran varias variables independientes se habla de modelo de Regresión Múltiple.

ANEXOS BIBLIOGRÁFICOS

Consideramos que los métodos estadísticos constituyen un valioso instrumento auxiliar de conocimiento. Por medio de esta metodología se consigue tanto la validez y representatividad formal de los criterios metodológicos a usar: diseño muestral de una población de estudio, extracción de conclusiones: pruebas de hipótesis, grados de asociación, etc...); además de aportar instrumentos o criterios para clasificar y ordenar datos.

_. Asociado del Centro Nacional de Consultoría.

_. Todo proceso cognitivo esta basado en una ideología personal o social conciente o inconsciente, explícita o implícita, que determina unas condiciones de selectividad tanto del tema como del proceso de identificación de un problema de investigación, al igual el camino metodológico a seguir y los criterios de análisis e interpretación.

_. No solo caracterizar -describir en términos cuantitativos sino también diferencias entre ellos, asociaciones o correlaciones.

_. Briones, Guillermo. M, todos y Técnicas Avanzadas de Investigación Aplicadas a la Educación y a las Ciencias Sociales. Módulo 4: Análisis e Interpretación. PIIE - ICFES. Bogotá. 1988. Pág. 21

_. Localización: se entiende como "el sitio" en donde los datos tienden a agruparse.

_. Briones. *Ibíd.* Pág. 23

_. Al 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.

_. Tablas extraídas del trabajo "La sexualidad en jóvenes adolescentes de sectores marginados." Universidad de Antioquia. Facultad de Ciencias Sociales y Humanas. Centro de Estudios de Opinión. 1992. Pág 27 y 28

_. La Moda puede ser utilizada también como medida de localización con datos ordinales e intervalos.

_. Diferente del "análisis de Varianza".

- _ . El Medellín que yo quiero. Universidad de Antioquia. Facultad de Ciencias Sociales. Depto. de Sociología. CEO. 1991. págs. 81
- _ . Briones. G. Op. cit. p. 56-57.
- _ . Índice de Movilidad Ocupacional = N° de ocupaciones desempeñadas ÷ N° de años trabajados * 10.
- _ . Briones. Guillermo. M, todos y técnicas de investigación para las Ciencias Sociales. Edit. Trillas. México. Pág178
- _ . Vélez, E. El Análisis de la información. Módulo 4 Serie Aprender a investigar. ICFES. Bogotá, 1987. Pág124
- _ . Briones. Op. cit. Pág56-59
- _ . La ejemplificación fue extraída del texto ya citado de Briones. Módulo 4. Op. Cit. Pág. 58 y 59
- _ . Las probabilidades asociadas con los valores F se computan mediante una distribución asociada como la "distribución F de Fisher -Snedecor.
- _ . Una distribución de frecuencia de datos que se parece gráficamente al corte transversal de una "campana" se le reconoce como una "distribución normal". La distribución normal es una distribución teórica y como es una distribución continua es necesario que los datos sean medidos en el nivel de intervalo o razón.
- _ . Martínez, J. et. Al. Módulo 3. Curso Especializado en Investigación en las Ciencias Sociales. ICFES - INER. 1992. Pág. 3.7
- _ . Ibíd. Pág3.8
- _ . Criterios de estratificación socio-económico de Planeación Nacional.
- _ . El valor 1.96 se encuentra en la Tabla de Distribución Normal.
- _ . No confundir la distribución de una muestra que son las características que tienen los datos a partir de una muestra, con la "distribución de muestreo."
- _ . La PRUEBA Z se utiliza cuando se conoce el parámetro de la población.
- _ . Ejemplificación extraída del Módulo 4. El análisis de la Información. Serie de Aprender a Investigar. ICFES. 1987. Pág. 119 -120

- _ . Solo en el momento en que la probabilidad es igual a uno (1), se tiene la certeza de que el suceso se va a realizar. Si un evento ocurre es porque su probabilidad de ocurrencia era alta.
- _ . El criterio de improbabilidad se lo denota por la letra griega ALFA (α) y comúnmente se le da un valor de $\alpha = 5\%$.
- _ . Briones. G. *Ibíd.* Pág. 76
- _ . En cuanto a la FORMA, es posible observar en algunos casos, una "relación de proporcionalidad directa con respecto a una variable, por ejemplo, al porcentaje de hogares con jefes analfabetos e "inversa" con respecto al porcentaje de hogares con jefe mujer y aparentemente ninguna con el porcentaje de hogares con jefes mayores de 65 años.