



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**INTRODUCCIÓN A LA ESTADÍSTICA: UN
ENFOQUE DIRIGIDO A LA SOCIOLOGÍA**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A:

GUILLERMO JAVIER CUEVAS PINEDA



**DIRECTOR DE TESIS:
MAT. MARGARITA ELVIRA CHÁVEZ CANO
2010**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Hoja de Datos del Jurado

<p>1. Datos del alumno. Autor. Apellido paterno: Apellido materno: Nombres Teléfono: Universidad Nacional Autónoma de México Facultad de Ciencias Carrera: No. de cuenta:</p>	<p>1. Datos del alumno. Cuevas Pineda Guillermo Javier 53 68 21 81 Universidad Nacional Autónoma de México Facultad de Ciencias Actuaría 098324239</p>
<p>2. Datos del tutor. Grado Nombres Apellido paterno: Apellido materno:</p>	<p>2. Datos del tutor Mat Margarita Elvira Chávez Cano</p>
<p>3. Datos del sinodal 1 Grado Nombre(s) Apellido paterno: Apellido materno:</p>	<p>3. Datos del sinodal 1 Dr José Rubén Hernández Cid</p>
<p>4. Datos del sinodal 2 Grado Nombre(s) Apellido paterno: Apellido materno:</p>	<p>4. Datos del sinodal 2 M en C Alejandro Mina Valdés</p>
<p>5. Datos del sinodal 3 Grado Nombre(s) Apellido paterno: Apellido materno:</p>	<p>5. Datos del sinodal 3 Dr Carlos Silva Forné</p>
<p>6. Datos del sinodal 4 Grado Nombre(s) Apellido paterno: Apellido materno:</p>	<p>6. Datos del sinodal 4 M en C Hugo Villaseñor Hernández</p>
<p>7. Datos del trabajo escrito. Título Subtítulo Número de páginas Año</p>	<p>7. Datos del trabajo escrito Introducción a la estadística: Un enfoque dirigido a la sociología 174 2010</p>

AGRADECIMIENTOS

Al Creador del Universo, por darme la oportunidad de vivir.

*A mi familia, principalmente a mis padres, porque sin su apoyo incondicional
habría sido imposible alcanzar esta meta.*

A mi directora de tesis, gracias por su paciencia, por su tiempo y sus consejos.

*A los sinodales de éste trabajo, por el tiempo y los gratos comentarios que le
dedicaron.*

*Un agradecimiento especial para la Fundación Alberto y Dolores Andrade, quienes
creyeron en mí, quienes fueron el aliciente y la guía a lo largo de mis estudios.*

*A todos los que me han acompañado en este largo trayecto, los que estuvieron,
los que están y los que estarán, gracias.*

TABLA DE CONTENIDO

0. Introducción	i
1. Introducción a la estadística descriptiva	1
1.1. La estadística en las ciencias sociales	1
1.2. Niveles de medición de las variables	2
1.3. Teoría, medición y matemáticas	5
1.4. El XII Censo Nacional de Población y Vivienda	9
1.5. Bibliografía del capítulo 1	12
2. Cuadros de doble entrada	14
2.1. Proporciones, razones, porcentajes	14
2.1.1. Definiciones	14
2.1.2. Aclaraciones	14
2.1.3. Ejemplos	15
2.2. Estructura, formato y lectura de datos	17
2.3. Construcción de indicadores demográficos y sociales	20
2.3.1. Reglas generales para la construcción de indicadores	21
2.4. Bibliografía	23
3. Conjuntos y números	24
3.1. Conjuntos y subconjuntos	24
3.1.1. El conjunto universo y el conjunto vacío	25
3.1.2. Subconjuntos	26
3.2. Operaciones con conjuntos	27
3.2.1. Unión	27
3.2.2. Intersección	30
3.2.3. El complemento	32
3.2.4. Diferencia	35
3.3. Una construcción conjuntista de los números	35
3.4. Igualdades y desigualdades	37
3.5. Bibliografía	40
4. Índices de precios y cantidad	41
4.1. Índice de precios	41
4.1.1. Índice de Laspeyres	41

4.1.2. El índice de Paasche	46
4.1.3. El índice de Fischer	52
4.2. Inflación, deflación, poder adquisitivo, valor del peso	54
4.2.1. Inflación	54
4.2.2. Deflación	56
4.2.3. Poder adquisitivo	56
4.2.4. Valor del peso	56
4.3. Cambio de base; empalme de índices	57
4.3.1. Cambio de base	57
4.3.2. Empalme de índices	58
4.4. Bibliografía	62
5. Estadística descriptiva	63
5.1. Introducción	63
5.2. Tabulación de los datos	63
5.2.1. Ejemplo 1	63
5.2.2. Ejemplo 2	65
5.2.3. Ejemplo 3	66
5.3. Medidas de posición	67
5.3.1. Datos sin agrupar	68
5.3.2. Datos agrupados	68
5.3.3. Los percentiles	69
5.3.3.1. Ejemplo para datos sin agrupar	70
5.3.3.2. Ejemplo para datos agrupados	74
5.3.4. Los deciles	77
5.3.4.1. Ejemplo para datos sin agrupar	78
5.3.4.2. Ejemplo para datos agrupados	85
5.3.5. Los cuartiles	87
5.3.5.1. Ejemplo	88
5.3.6. Conclusión	91
5.4. Medidas de tendencia central	91
5.4.1. La media	92
5.4.1.1. Ejemplo 1	92
5.4.1.2. Ejemplo 2	93
5.4.1.3. Ejemplo 3	94

5.4.1.4.	Ejemplo 4	94
5.4.1.5.	Ejemplo 5	95
5.4.2.	La moda	97
5.4.2.1.	Ejemplo 1	97
5.4.2.2.	Ejemplo 2	98
5.4.3.	La mediana	98
5.4.3.1.	Ejemplo 1	99
5.4.3.2.	Ejemplo 2	99
5.4.3.3.	Ejemplo 3	100
5.4.3.4.	Ejemplo 4	100
5.4.3.5.	Ejemplo 5	101
5.4.3.6.	Conclusión	101
5.5.	Medidas de dispersión	102
5.5.1.	La amplitud	102
5.5.1.1.	Ejemplo 1	102
5.5.1.2.	Ejemplo 2	104
5.5.2.	La amplitud intercuartil	105
5.5.2.1.	Ejemplo	105
5.5.3.	La amplitud semiintercuartil	106
5.5.3.1.	Ejemplo 1	107
5.5.3.2.	Ejemplo 2	107
5.5.3.3.	Ejemplo 3	108
5.5.3.4.	Conclusión	109
5.5.4.	La desviación media	109
5.5.4.1.	Fórmula para datos sin agrupar	110
5.5.4.2.	Fórmula para datos agrupados	111
5.5.4.3.	Ejemplo 1	111
5.5.4.4.	Ejemplo 2	112
5.5.4.5.	Ejemplo 3	114
5.5.4.6.	Conclusiones	115
5.5.5.	La desviación estándar	115
5.5.5.1.	Ejemplo 1	116
5.5.5.2.	Ejemplo 2	117
5.5.5.3.	Ejemplo 3	119
5.5.5.4.	Ejemplo 4	120
5.5.5.5.	Ejemplo 5	122

5.5.6. La varianza	123
5.5.7. El coeficiente de variación	123
5.5.7.1. Ejemplo 1	124
5.5.7.2. Ejemplo 2	125
5.5.8. Un ejemplo general	126
5.6. Representación gráfica	128
5.6.1. El histograma	127
5.6.1.1. Ejemplo 1	128
5.6.1.2. Ejemplo 2	130
5.6.2. El polígono de frecuencias	132
5.6.2.1. Ejemplo	133
5.6.3. La ojiva	136
5.6.3.1. Ejemplo	137
5.7. La sumatoria	139
5.8. Bibliografía	141
6. Medidas de concentración	142
6.1. Introducción	142
6.2. La interpolación polinomial	143
6.2.1. Los polinomios	143
6.2.2. El algoritmo de interpolación	145
6.2.2.1. Ejemplo 1	146
6.2.2.2. Ejemplo 2	147
6.3. El área de un polinomio entre 0 y 1	149
6.4. La curva de Lorenz	150
6.5. El índice de Gini	150
6.6. Ejemplo 1	150
6.7. Ejemplo 2	154
6.8. Conclusión	157
6.9. Bibliografía	158
7. Censos Nacionales y Estadísticas Continuas	159
7.1. Censo General de Población y Vivienda	159
7.2. Censo Económico	161
7.3. Censo Agrícola y Ganadero	163

7.4. Anuario Estadístico de la UNAM	164
7.5. Anuario Estadístico del INEGI	165
7.6. Anuario Estadístico de la ONU	167
7.7. Registro Nacional de Población	168
7.8. Bibliografía	169
8. Conclusiones Generales	170
9. Bibliografía	173

Introducción

En la actualidad, el aprendizaje y uso de la estadística es requerido en prácticamente todas las ciencias sociales. En ese sentido, los estudiantes de sociología tienen la necesidad de aprender a utilizar las herramientas que la estadística provee. Para ello, en el aprendizaje que realizan en las universidades, y fuera de ellas, recurren a la bibliografía existente. Desafortunadamente, la mayoría de los textos están escritos para otro tipo de público (economistas, ingenieros); están escritos en otro idioma; están desactualizados; o bien, presentan ejemplos concernientes a otros países y realidades que en principio parecen ajenas a la mexicana. Estoy convencido de que el aprendizaje de la estadística se facilita en la medida en que se eviten las desventajas ya enumeradas. Así, éste trabajo presenta una propuesta de libro de texto sobre estadística para estudiantes de licenciatura de la carrera de sociología de la Facultad de Ciencias Políticas y Sociales de la UNAM.

La Facultad de Ciencias Políticas y Sociales admite a aproximadamente 200 alumnos anualmente para cursar la carrera de sociología, y considerando a los 800 alumnos de reingreso, se suma una población de 1,000 alumnos que deberán aprobar, al menos, cuatro materias relacionadas con la estadística, a saber: estadística descriptiva; estadística inferencial; análisis cuantitativo; y procesamiento de datos. Esta tesis desarrolla lo establecido en el plan de estudios de la carrera para la materia de “Estadística Descriptiva”, y como el mismo plan menciona, pretende coadyuvar para que el alumno que cursa la materia “tenga los conocimientos y la capacidad de recopilar, tabular, calcular, analizar e interpretar la información estadística a nivel descriptivo y comparativo”.

La tesis está estructurada en 7 capítulos que desbastan la mayor parte del material previsto por el plan de estudios de la materia. El primer capítulo presenta una breve introducción tanto a la pertinencia de la materia como a los principales conceptos que serán utilizados a lo largo del texto. El segundo capítulo maneja el tema de cuadros de doble entrada, la forma de presentación de los datos, y los criterios mayormente

utilizados en la misma. Durante el tercer capítulo se expone de manera general el concepto de conjunto y sus principales operaciones, dicho concepto será de utilidad para el alumno al momento de estudiar estadística inferencial; el capítulo finaliza con una breve exposición acerca de la construcción conjuntista de los números. Estos antecedentes permiten introducir de manera más clara, en el capítulo cuarto, el concepto de índices de precios y de cantidades. El capítulo se concentra en presentar al Índice de Laspeyres, al Índice de Paasche y al Índice de Fisher, su comparación, y el cambio de base en un mismo índice, pero también presenta algunos conceptos básicos de economía, como son la inflación, la deflación, el poder adquisitivo, entre otros.

El quinto capítulo de la tesis representa la parte medular de la misma: la estadística descriptiva. En él, se desarrollan conceptos como la tabulación de los datos, las medidas de posición (los percentiles, los deciles, y los cuartiles), las medidas de tendencia central (la media, la moda y la mediana), las medidas de dispersión (la amplitud, la amplitud intercuartil, la amplitud semi-intercuartil, la desviación media, la desviación estándar, la varianza, y el coeficiente de variación), y la representación gráfica de dichas medidas (el histograma, el polígono de frecuencias y la ojiva). Los conceptos presentados se ejemplifican con datos provenientes de encuestas y censos mexicanos recientes, que al ejemplificar el uso de las distintas medidas y fórmulas, también ejemplifican la pertinencia de su conocimiento.

El sexto capítulo abarca el tema de medidas de concentración, describiendo para ello el uso de la curva de Lorenz y el índice de Gini. Para ello, se desarrolla el tema de la interpolación polinomial. Cabe destacar que estos conceptos se aplican tanto a datos monetarios como a datos de otra índole, ejemplificando con ello que el rango de aplicaciones puede ser mayor al rango de aplicaciones comunes. Finalmente, en el séptimo capítulo se mencionan algunos datos de los censos nacionales (como son el de población y vivienda, el económico y el agrícola y ganadero), y las estadísticas continuas (en donde se mencionan al anuario estadístico de la UNAM, el anuario estadístico del INEGI, y el anuario estadístico de la ONU), así como la utilidad que éstos datos tienen.

Con los temas así desarrollados, se abarca casi en su totalidad el temario descrito para la materia Estadística Descriptiva. Sin embargo, hubo dos temas que, estando en el temario de la materia, no se desarrollaron en el presente texto: El álgebra de números (donde se abarcan temas como la suma, multiplicación y división de expresiones algebraicas; los productos especiales; los factores y descomposición de factores; la simplificación, suma, multiplicación y división de fracciones; los exponentes enteros y racionales; y los radicales), así como el tema de series históricas (en donde se habrían desarrollado temas como funciones lineales, logarítmicas y exponenciales; exponentes y logaritmos comunes; progresiones geométricas; interés compuesto; y, gráficas semilogarítmicas). Estos temas no fueron desarrollados porque, además de existir una amplia bibliografía donde son desarrollados con suficiente claridad, son temas que se encuentran dentro de los planes y programas de estudio de nivel medio y medio superior. Además, son temas que no serán necesarios, ni para el presente texto, ni para las materias consecuentes a ésta.

Finalmente, quisiera expresar mi gratitud al Área de Investigación Aplicada y Opinión del Instituto de Investigaciones Jurídicas por la facilidad de acceso a las diversas bases de datos provenientes de encuestas, que son utilizadas en la mayoría de los ejemplos de éste trabajo. En particular, sea mi reconocimiento a su coordinadora general, la Mtra. Julia Isabel Flores Dávila, quien ha hecho todo esto posible.

I.1 INTRODUCCION A LA ESTADÍSTICA DESCRIPTIVA

1.1 La estadística en las ciencias sociales.

Uno de los objetivos principales de las ciencias sociales y del comportamiento es desarrollar principios que expliquen y provean nuevas formas de conocer y comprender el comportamiento humano. La estadística juega un papel vital para recabar, resumir e interpretar datos recopilados constantemente por los seres humanos para tratar de comprender su entorno. Consecuentemente, el entendimiento de esta materia es necesario para controlar la investigación y para realizar una evaluación crítica de las investigaciones realizadas.

El campo de la estadística tiene aplicaciones cada vez más amplias en diversas disciplinas, en la ingeniería, la zoología, la sociología, la economía, etcétera. Por lo anterior, la estadística es una ciencia en constante desarrollo, para poder atender a las diversas necesidades planteadas por investigadores tan diversos. Podría suponerse por ello, que la estadística es una ciencia muy compleja, o demasiado amplia para su comprensión. La abstracción, es la que consigue que esto no sea así. Por medio de la abstracción, datos tan diversos como los provenientes de arquitectos y de politólogos pueden ser resumidos siguiendo los mismos lineamientos generales.

La estadística descriptiva tiene como su principal función la organización de los datos y el cálculo de índices numéricos¹ para una muestra dada².

El papel de la estadística en el proceso de investigación sociológica está claramente determinado. La estadística se utiliza para operar con números, que reflejan valores de mediciones que se supone satisfacen ciertos supuestos. Esto es, las consideraciones estadísticas se introducen tan sólo en la fase

¹ Un índice numérico es un número que entre más grande o más pequeño sea, representa que una cualidad está más o menos presente dentro del dato. El concepto de índice implica por lo regular que el procedimiento empleado da sólo un indicador imperfecto de alguna variable señalada que no es medible directamente (Blalock 1966, pp 21-22)

² Una muestra es una parte de nuestra población objetivo. Por ejemplo, en biología, si queremos estudiar a los conejos por medio de ciertas mediciones, en lugar de trabajar con todos los conejos del mundo, se selecciona a 10, y en este caso, decimos que tenemos una muestra de tamaño 10.

analítica del proceso de investigación, después de haber obtenido los datos frecuentemente a partir de una muestra. Debe, pues, quedar claro desde el principio, que la estadística es siempre una buena ayuda, pero nunca un sustituto para un buen razonamiento teórico, y un buen quehacer metodológico. En la historia de la sociología como ciencia, el desarrollo del proceso de cuantificación ha jugado un papel ciertamente importante en la configuración de la sociología como una disciplina científica. Los fundadores de la sociología, Marx, Durkheim, Weber, Quetelet, Le Play, entre muchos otros, reconocieron la importancia de la obtención de información cuantitativa relevante sobre los fenómenos sociales, y de su tratamiento estadístico, para construir una ciencia sobre la sociedad. Durkheim fue de los primeros en reconocer el fundamental papel de la estadística para aislar apropiadamente los hechos sociales de los hechos particulares³.

Es entonces, por medio de la estadística, por el que se pueden comprobar o no ciertas características de la sociedad, en particular características sociales tan diversas como son, pobreza, marginación, preferencia política, situaciones de riesgo, enfermedad, mortalidad, natalidad, etcétera. Con el auge que el mundo actual otorga al análisis cuantitativo, poco a poco, nos hemos llenado de información, que en muchos casos puede ser manejada de tal forma que resulte engañosa. Por esta razón, es muy importante realizar el procesamiento adecuado de los datos, siguiendo métodos generales, o sea, utilizando adecuadamente a la estadística descriptiva.

1.2 Niveles de medición de las variables.

En adelante, entenderemos como variables a todas las características que se medirán de la muestra o de la población, y que por su naturaleza son cambiantes de individuo a individuo. Por ejemplo, de un grupo de estudiantes, pueden ser variables el sexo, el número de materias aprobadas, la estatura, el año de nacimiento, su nivel de ingresos, su estado civil, si sufren de violencia intrafamiliar, etcétera.

³ Véase García, 1999, pág. 29

Para que una variable pueda ser susceptible de análisis estadístico, necesita poder medirse o cuantificarse de alguna forma. La medición consiste, a grandes rasgos, en asignar números a las observaciones de acuerdo a ciertas reglas. El tipo de procesamiento estadístico depende, en parte, del tipo de medición que se hizo de la variable, por lo cual es importante estudiar los diferentes tipos de medición de las variables. Los siguientes son los niveles de medición de variables, desde el punto de vista estadístico:

- Nivel Nominal: Este es el nivel más simple de medición. Los números son simplemente usados para identificar o nombrar el atributo o categoría descrito. Por ejemplo, si nuestra variable es el sexo, dicha variable es nominal si asignamos el número 1 a los hombres y el número 2 a las mujeres (en realidad cualquier pareja de números distintos podría ser útil con estos fines). Si la variable se refiere a si sufren de violencia intrafamiliar, se podría asignar el número 0 a los que no lo padecen y el número 1 a los que sí lo padecen. A estos datos a veces se les dice categóricos o de frecuencia, y como los números sólo sirven para identificar una cualidad, no tiene sentido realizar operaciones aritméticas con los mismos (suma, resta, multiplicación, división u ordenamiento).

- Nivel Ordinal: Es un nivel más complejo que el anterior. En él, los números indican una cantidad baja, o alta, de un atributo en particular. Cuando se utiliza este tipo de medición, los datos deberán estar acomodados de forma creciente, o decreciente, según sea el caso. Por ejemplo, al número de materias aprobadas en un semestre dado, se pueden asignar: 0 a los que tengan 0 materias aprobadas, 1 a los que tengan 1 materia aprobada, y así sucesivamente, y entonces seguiremos un orden creciente (en cuanto a número de materias aprobadas). También es posible, siguiendo con el mismo ejemplo, que asignemos el 0 a los que tengan 8 materias aprobadas (suponiendo que 8 es el número máximo), 1 a los que tengan 7 materias aprobadas, y así sucesivamente, y en este caso se seguirá un orden decreciente. Por último, hay que señalar que no se tienen que utilizar estos números específicamente, si en lugar de utilizar 0, 1, 2, ..., 8, se utilizan 2, 34, 67, 3000, 28930, 67750, 100293, 328900, 576930, no habría mayor

complicación, desde el punto de vista estadístico, ya que se sigue manteniendo un orden. Nótese, además, que una variable de tipo ordinal, cumple con las propiedades de una variable de tipo nominal, esto es, los números que se usan diferencian perfectamente a las categorías, pero además existe un orden, lo que no ocurre con una variable que sólo es nominal (en general, no se puede decir que el hombre sea más o menos importante que la mujer, o que tenga alguna precedencia).

- Nivel de Intervalo: En este nivel, los intervalos entre números denotan cantidades iguales del atributo en estudio, pero no se tiene un punto de partida o cero. Por ejemplo, entre una persona que nació en 1900 y otra que nació en 1905 hay el mismo número de años (5) que entre una persona que nació en 1998 y otra que nació en 2003. Sin embargo, como las escalas de intervalo no tienen un cero absoluto (o punto de partida), no se pueden expresar cantidades dentro de ellas (como que un valor es del doble que el otro). Por ejemplo, no podemos decir que una persona que nació en 1900 tiene un 5% menos de edad que una persona que nació en 2000, y esto debido a que no existe un punto de partida (puesto que pueden existir mediciones menores a 1900). Entre un fósil que data del año 500 y un fósil que data del año 9000 AC ni siquiera habría un punto de comparación porque se atraviesa el año cero. Si bien es posible decir que el fósil del año 500 es 9500 años más joven que el anterior (y este tipo de comparaciones siempre serán posibles con variables de intervalo), no se puede decir que ése mismo fósil sea 3, 30 o 100 veces más joven que el anterior. Por último, obsérvese que una variable de intervalo es variable de razón al mismo tiempo.

- Nivel de Razón: Representa al nivel más sofisticado de medición y tiene todas las propiedades de los anteriores, además incluye un cero absoluto, o punto de partida. Regresando al caso del año de nacimiento, si suponemos que nuestro cero absoluto es 1900 (es decir, no hay nadie por debajo de esa medición), una persona que nació en 1950 y otra que nació en 2000 se puede decir que el primero es el doble de viejo que el segundo. Un ejemplo más claro se puede apreciar en el caso de las estaturas, si alguien mide 1 metro y otro estudiante mide 2 metros, se dice que el segundo es el doble de alto que el primero. Sin embargo, si ponemos al cero absoluto o punto de partida en 1.25

metros, se puede decir que si un estudiante mide 1.50 metros y el otro 2.00 metros, el segundo mide 3 veces más que el primero.

Hay que destacar que siempre que se tenga una variable de tipo nominal, se tiene que estudiar para ver si puede cumplir con los requisitos para ser una variable ordinal. Si los cumple, habrá que ver si tiene las características para que sea variable de intervalo. Si las tiene, se verificará si puede ser variable de razón. Lo anterior es muy importante, porque para una variable, entre mayor sea su nivel de sofisticación, será susceptible de un análisis estadístico más complejo que pueda derivar en conclusiones más interesantes, o de mayor relevancia. Para terminar, el lector seguramente habrá notado que siempre que se esté en un nivel superior se cumplen todas las características de todos los niveles inferiores.

1.3 Teoría, medición y matemáticas⁴.

Pese a que en la actualidad en muchos casos se trabajan por separado, la teoría, los métodos de investigación y la estadística de las materias están muy relacionadas entre sí. Con objeto de situar a la estadística en la debida perspectiva, conviene prestar atención a las relaciones entre las proposiciones teóricas y las hipótesis de investigación por una parte, y entre estas últimas y los modelos matemáticos por la otra.

Se suele decir con frecuencia que el objeto de la investigación está en verificar hipótesis desarrolladas teóricamente y que los métodos estadísticos capacitan para efectuar dichas pruebas. Hay que tener presente, sin embargo, que los procesos implicados en pasar de la teoría a las hipótesis reales de investigación y de éstas a la estadística o a la probabilidad⁵ no son en modo alguno directos.

En el instante en que empezamos a diseñar un proyecto de investigación dirigido a verificar una proposición que puede aparecer en un trabajo teórico,

⁴ Véase Blalock, 1966, pp 20 a 33.

⁵ La probabilidad es la rama de las matemáticas en la que la estadística basa gran parte de sus desarrollos y supuestos. Dicha rama es tratada en el capítulo II-2.

resulta evidente que hay que hacer varias cosas antes de proceder a la prueba. Tomemos como ejemplo a la siguiente proposición: “Cuanto más elevada es la condición social de una persona, tanto menores son sus prejuicios en relación con los indígenas”. Supongamos que la “condición social” se haya definido como la posición que la persona ocupa en relación con otras en la jerarquía social, y los “prejuicios” como tendencia latente a la discriminación de una minoría o como actitud negativa basada en juicios preformados. Aun si se prefiere sustituir por otras definiciones de esos dos conceptos, se descubrirá sin duda alguna que, cualesquiera que sean las definiciones escogidas, resulta imposible servirse directamente de ellas para decidir cuál es el grado de sus prejuicios.

La razón de ello reside en que la mayoría de las proposiciones son más bien teóricas que operativas. En la definición teórica, en efecto, un concepto se define en términos de otros conceptos que se dan por comprendidos. En el modelo ideal del sistema completamente deductivo, se tomarían ciertos conceptos sin definir, y todos los demás se definirían en términos de aquellos⁶. Pese a que la elección de conceptos no definidos es hasta cierto punto arbitraria, el hecho de que tenga que haber siempre algunos conceptos primeros o primarios resulta de la necesidad inherente de definir los conceptos teóricos en términos unos de otros.

Por otra parte, las definiciones operativas son definiciones que enuncian efectivamente los procedimientos empleados en la medición. El ejemplo de una definición operativa del prejuicio implicará una prueba como la de la escala de la distancia social de Bogardus⁷. El índice de las características sociales de Warner⁸ constituye un ejemplo del procedimiento empleado al formular definiciones operativas de la condición social. Como quiera que toda medición implica como requisito mínimo alguna clasificación, la definición operativa puede considerarse como un conjunto detallado de instrucciones que permiten

⁶ Como un ejemplo, Kronecker propuso a fines del siglo XIX una serie de proposiciones que se tomarían por ciertas (en adelante axiomas), y en base a las cuales se tendría que construir cualquier otro conocimiento en matemáticas.

⁷ La escala social de Bogardus mide la cantidad de contactos que voluntariamente una persona quisiera tener con diversos grupos sociales, incluyendo aquellos con los que no se identifica (véase Babbie, 2004).

⁸ El índice de características sociales de Warner, creado en 1949 y aún en uso (con ciertas modificaciones), mide el estatus socioeconómico de una persona a través de su ocupación, su nivel de ingresos, su domicilio y la colonia o barrio en el que habita (véase Werner, 1949).

clasificar a los individuos en forma inequívoca. La definición ha de ser lo bastante precisa para que todas las personas que se sirvan del procedimiento lleguen a los mismos resultados. Lo que las definiciones teóricas del prejuicio y de la condición social no lo permitirían directamente.

La gente, en general, está de acuerdo en que debería emplearse una determinada operación operativa como medida de un determinado concepto, si las operaciones parecen razonables sobre la base de la definición teórica. Puede presumirse que, si varias definiciones operativas son posibles, se escogerán aquellas que parecen más apropiadas y al mismo tiempo más seguras. El carácter de “apropiado” ha de juzgarse inevitablemente sobre la base de la comprensión que uno tiene de la definición teórica. Se emplea a veces el término de ‘validez aparente’ para designar el carácter apropiado de un índice o de una definición operacional. Idealmente, las operaciones y las definiciones teóricas habrían de asociarse sobre la base de uno a uno.

Podemos admitir la posibilidad de tener asociado un número de diversas operaciones o de índices a cada concepto teórico. Pero en tal caso podemos encontrarnos con una dificultad común: dichos procedimientos pueden dar resultados distintos. Uno de los procedimientos empleados para medir el prejuicio puede llevar a resultados que indiquen que nuestra hipótesis ha sido confirmada. En tanto que, en otro caso, otro procedimiento puede conducir a la conclusión opuesta. En cierto sentido, así se depuran las mediciones cuantitativas, a condición de que no conduzcan a una disputa interminable acerca de cuál es el procedimiento que realmente mide al prejuicio. Con objeto de prevenir confusiones importa darse cuenta de que la prueba efectiva se hace en términos de los conceptos tal como se los ha definido operacionalmente. Por lo tanto, las proposiciones que comportan conceptos definidos teóricamente no son verificables directamente. Así, pues, si se dan dos definiciones operativas distintas del prejuicio se verificarán dos hipótesis distintas.

Ahora bien, supongamos que tenemos un concepto definido teóricamente y varias definiciones operativas susceptibles de asociarse con

dicha definición teórica. Según lo ya expuesto, la mayoría de los científicos estarán probablemente de acuerdo en que algunas de las operaciones deberían eliminarse por cuanto no se aplican a lo que está contenido en la definición teórica. Pueden decidir, por ejemplo, que las preguntas relativas a las tendencias delictivas o los gustos musicales no deberían emplearse para medir el prejuicio. Pero puede haber varias operaciones que ocupen más o menos el mismo lugar en la opinión de dichos jueces. En otros términos: sobre la base de la definición teórica, los expertos pueden no estar en condiciones de ponerse de acuerdo acerca de que un determinado procedimiento operativo debería escogerse en preferencia a otros. Podemos decir en tal caso que, en la medida en que dichos diversos procedimientos dan resultados diferentes, la definición teórica es deficiente, en el sentido de que necesita probablemente revisión o aclaración. Por ejemplo: el concepto prejuicio puede acaso haberse definido de tal modo que resulte demasiado vago. Tal vez se considerará necesario distinguir entre varias clases o dimensiones del prejuicio, asociando operaciones distintas a cada una de ellas. En estos casos, el proceso de investigación puede utilizarse para ayudar a aclarar los conceptos teóricos.

En resumen, los científicos, en general, piensan en el lenguaje teórico y realizan sus experimentos en el lenguaje operativo. No es necesario asociar operaciones con todos los conceptos del lenguaje teórico. Sin embargo, importa percatarse de que los conceptos que no han sido definidos operativamente no deberían por lo regular aparecer en enunciados que pretenden constituir hipótesis comprobadas.

Otro de los problemas que se presentan entre ambos lenguajes es que el lenguaje operativo comporta determinadas operaciones físicas, tales como el empleo operativo comporta determinadas operaciones físicas, tales como el empleo de una unidad de medida. El lenguaje matemático, a su vez, implica un sistema totalmente abstracto de símbolos y operaciones matemáticas, y es útil no sólo porque es preciso y está altamente desarrollado, sino debido también a que su carácter abstracto, lo que permite la aplicación a una gran variedad de problemas empíricos. Las matemáticas se sirven del razonamiento deductivo por el que se pasa de un conjunto de definiciones, supuestos y reglas de

operación a un conjunto de conclusiones mediante un razonamiento puramente lógico. En sí mismas, las matemáticas nada nos dicen acerca de la realidad, ya que todas las conclusiones están contenidas en las definiciones, los supuestos y las reglas originales, no habiéndose determinado empíricamente. Así, pues, si han de ser de alguna utilidad para el científico las conclusiones matemáticas han de traducirse inversamente a los lenguajes operativo y teórico. Por ello, no es legítimo, por ejemplo, servirse de un sistema matemático que comporta las operaciones de sumar o restar, si esto no está legitimado por el método de medición, en términos más generales, no podemos remontar legítimamente en la jerarquía de medición, a menos que el proceso mismo de la medición haya sido mejorado.

Para terminar, seamos más enfáticos. Toda técnica estadística considerada presupone siempre un nivel específico de medición requerido por cada procedimiento. La práctica actual, ha hecho que las técnicas estadísticas de vanguardia se desarrollen alrededor de los niveles más bajos, debido a la necesidad imperiosa de los científicos de éstas herramientas, ya que en muchos casos alcanzar niveles de intervalo o de razón era prácticamente imposible dentro del lenguaje operativo. Volvemos a repetir: al servirse de una técnica estadística cualquiera, hay que darse cuenta cabal de los supuestos importantes que hay que adoptar.

1.4 El XII Censo Nacional de Población y Vivienda

En ésta sección nos dedicaremos únicamente a la revisión del cuestionario básico del XII Censo de Población y Vivienda⁹. A manera de ejemplo, veremos los diversos niveles de las variables, o preguntas, que maneja el cuestionario.

Sección I

Pregunta 1. En este caso, que se refiere al material de las paredes de la vivienda, el tipo de escala es nominal. Así, el 1 representa a

⁹ Véase Anexo I

“Material de desecho”, el 2 “Lámina de cartón”, y así sucesivamente, sin que en realidad los números representen algún tipo de orden.

Pregunta 2. Igual que en el caso anterior, variable es del tipo nominal.

Pregunta 3. En este caso, también la escala es de tipo nominal.

Pregunta 4. Las dos partes en que se dividió ésta pregunta también corresponden a escalas de nivel nominal.

Pregunta 5. En éste caso, pese a que la variable (en ambos casos) sólo toma valores enteros (es decir, toma valores como 0, 1, 2, 3, ..., pero no toma valores como 1.5, 3.54, u otros), la variable tiene el nivel de razón, ya que cuenta con un cero absoluto (el 0, nadie puede tener menos de 0 cuartos), y con los números se pueden realizar todas las operaciones matemáticas necesarias (si por ejemplo, alguna vivienda tiene 3 cuartos para dormir sin contar pasillos, y tiene en total 10 cuartos, sin contar pasillos y contando la cocina, se puede decir que de esos 10 cuartos, 7 ($10 - 3$) son usados para fines distintos del de dormir).

Pregunta 6. Cada uno de los atributos (los casos) corresponde a una variable de nivel nominal, ya que se cuenta con ese servicio o no se cuenta con él.

Pregunta 7. Igual que en caso anterior, la variable es de nivel nominal.

Pregunta 8. En éste caso, también la variable es de nivel nominal.

Pregunta 9. También en éste caso la variable es de nivel nominal.

Pregunta 10. Variable de nivel nominal.

Pregunta 11. Variable de nivel nominal.

Pregunta 12. Variable de nivel nominal.

Pregunta 13. Variable de nivel nominal.

Pregunta 14. Variable de nivel nominal.

Sección II

Pregunta 1. Igual que en caso de la pregunta 5 de la sección I, pese a que la variable toma únicamente valores enteros, la variable es de nivel de razón.

Pregunta 2. Variable de nivel nominal.

Pregunta 3. Como ya se ha visto, en vista de la existencia de un cero absoluto, la variable es de nivel de razón.

Sección III

Pregunta 1. Variable de nivel nominal.

Pregunta 2. Variable de nivel nominal.

Pregunta 3. Debido a la existencia de un cero absoluto, ésta variable es de nivel de razón.

Pregunta 4. Variable de nivel nominal.

Pregunta 5. Variable de nivel nominal.

Pregunta 6. Variable de nivel nominal.

Pregunta 7. Variable de nivel nominal.

Pregunta 8. Variable de nivel nominal.

Pregunta 9. Variable de nivel nominal.

Pregunta 10. Variable de nivel nominal.

Pregunta 11. Variable de nivel nominal.

Pregunta 12. En éste caso, se podría decir que la variable es de nivel nominal, sin embargo, resulta claro que las opciones de respuesta han sido colocadas de tal manera que es clara la existencia de un orden, razón por la cual se dice que la variable es de nivel ordinal.

Pregunta 13. Variable de nivel nominal.

Pregunta 14. Variable de nivel nominal.

Pregunta 15. Variable de nivel nominal.

Pregunta 16. Variable de nivel nominal.

Pregunta 17. Variable de nivel nominal.

Pregunta 18. Variable de nivel nominal.

Pregunta 19. Variable de nivel nominal.

Pregunta 20. Variable de nivel nominal.

Pregunta 21. Nuevamente, en éste caso existe un cero absoluto, por ello, se dice que la variable es de nivel de razón.

Pregunta 22. Igual que en caso anterior, el ingreso nunca puede ser negativo, por lo cual existe un cero absoluto, y se dice que la variable es de nivel de razón. En cuanto al periodo de pago, se puede ver que las opciones han sido ordenadas de mayor periodicidad a menor periodicidad por lo cual ese caso es una variable de nivel ordinal.

Pregunta 23. Variable de nivel nominal.

Pregunta 24. Como ya se ha visto en otros casos, ésta variable es de nivel de razón.

Pregunta 25. Variable de nivel de razón.

Pregunta 26. Variable de nivel de razón.

Pregunta 27. El mes corresponde a una variable de nivel ordinal, ya que en realidad, por ejemplo, septiembre no es mayor a junio. En cuanto al año, y tal y como habíamos visto ya en el ejemplo, debido a la inexistencia de un cero absoluto corresponde a una variable de nivel de intervalo.

Pregunta 28. Variable de nivel nominal.

Pregunta 29. En todos los casos (días, meses o años), la variable es de nivel de razón.

Como se puede apreciar, la mayor parte de las preguntas son del tipo nominal (pese a que algunas podrían ser consideradas como variables de nivel ordinal). También existen variables de nivel ordinal y de razón, y una variable de nivel de intervalo, y, tal y como se había expuesto, las variables de nivel de intervalo son poco comunes. En capítulos posteriores se analizarán una serie de datos estadísticos, provenientes de órganos gubernamentales nacionales (como el INEGI), así como de otras instancias nacionales recopiladoras de datos (principalmente a través de encuestas), y en algunos casos se utilizarán datos internacionales, aunque siempre se hará un énfasis en los datos mexicanos.

1.5 Bibliografía.

[1] **García Ferrando**, Manuel. *Socioestadística Introducción a la Estadística en Sociología*. Alianza Editorial Manuales / Ciencias Sociales. España, 1999.

[2] **Babbie**, Earl. *The practice of social research*. Belmont, CA. USA, 10ª edición, 2004.

[3] **Blalock**, Hubert M. Jr. *Estadística Social*. Fondo de Cultura Económica. México, 1966.

[4] **Cramer**, Duncan. *Basic statistics for social research*. Routledge London and New York. UK, 1997.

[5] **Garrido Luque**, Alicia, et. al. *Técnicas de Análisis Estadístico en Ciencias Sociales*. Servicio de Publicaciones Universidad Complutense. España, 1995.

[6] **Croxtan**, Frederick E. *Estadística General Aplicada*. Fondo de Cultura Económica. México, 1948.

[7] **Nidditch**, P. H. *El desarrollo de la Lógica matemática*. Ediciones Cátedra, Col. Teorema, 5ª edición. España, 1995.

[8] **Werner**, Lloyd. *Social Class in America*. Science research associates, inc, Chicago, USA, 1949.

I.2. CUADROS DE DOBLE ENTRADA

Cuando se han recolectado datos de diversas fuentes, para un manejo más eficiente de los mismos es necesario resumirlos y hacerlos comparables. Las medidas que se examinan en el presente capítulo permiten establecer comparaciones entre diversos grupos, mediante normalización, esencialmente en relación con el tamaño.

2.1 Proporciones, razones, porcentajes.

2.1.1 Definiciones

Proporción: Se define la proporción de una clase A respecto a otra B , como al número de elementos en la clase de A entre el número de elementos en la clase de B (regularmente, la clase B contiene a la clase A , y tiene más elementos). La proporción de casos en cualquier categoría dada está definida como el número en la categoría dividido entre el número total de casos.

Razón: Si tenemos dos números, A y B , la razón de A a B se define como A/B (La división de A entre B).

Porcentaje: Cuando una clase B contiene a la clase A y a todos sus elementos, entonces, el porcentaje se define como la proporción de A a B , multiplicando el número resultante por 100.

2.1.2 Aclaraciones

Con objeto de podernos servir de las proporciones, hemos de presumir que el método de clasificación ha sido tal que las categorías son mutuamente excluyentes y exhaustivas¹. Lo anterior significa que para fines de comparabilidad y de un mejor manejo de los datos, siempre debe existir una

¹ Se refiere a que cada elemento está en una clase y sólo en una.

clase que sea la mayor (se podría ver como el conjunto del que sale todo), y las proporciones se deben hacer con respecto a ese conjunto total.

Los porcentajes pueden obtenerse de las proporciones multiplicando simplemente por 100. La palabra porcentaje significa por cien. Por lo tanto, al servirnos de los porcentajes normalizamos en relación con el volumen, calculando el número de individuos que habría en una categoría determinada si el total de los casos fuera 100, permaneciendo inalterada la proporción en cada categoría.

Como ya se dijo, la razón de un número A con respecto a otro número B se define como A dividido entre B . La cantidad que precede se pone en el numerador, en tanto que la que sigue forma el denominador. Obsérvese que, a diferencia de la proporción, la razón puede tomar un valor superior a la unidad. Por ello, se dice que las proporciones representan un tipo especial de razón en la que el denominador es el número total de casos y el numerador una cierta fracción de aquel. Sin embargo, el término de razón se emplea por lo regular para referirse a casos en los que A y B representan categorías separadas y distintas.

2.1.3 Ejemplos

A la pregunta: '¿La democracia en México es hoy: muy estable, bastante estable, poco estable o nada estable?'², se contestó como sigue:

CUADRO I.2-1
¿La democracia en México es hoy...?

Nada estable	24
Poco estable	207
Bastante estable	220
Muy estable	49

Introducción a la estadística: Un enfoque dirigido a la sociología.

FUENTE: Equipo de investigación sobre élites parlamentarias, UDES0, IISUNAM. Estudio levantado entre los 500 diputados de México en Abril/Julio 2004.

² FUENTE: Equipo de investigación sobre élites parlamentarias, UDES0, IISUNAM. Estudio levantado entre los 500 diputados de México en Abril/Julio 2004.

La proporción de ‘Nada estable’ es el número de personas que contestaron esa respuesta, entre el número total de personas que participaron, esto es: $24/(24+207+220+49) = 24/500 = 12/250 = 6/125^3$.

El porcentaje de ‘Poco estable’ es el número de personas que contestaron a esa respuesta, entre el número total de personas que participaron, y esto por 100, esto es: $[207/(24+207+220+49)]*100 = (207/500)*100 = (0.414)*100 = 41.4\%^4$.

La razón de ‘Muy estable’ a ‘Nada estable’ es el número de personas que contestaron la primera respuesta, entre las que contestaron la segunda, es decir: $49/24 = 49:24^5 = 2.04 = 2.04:1$.

A la pregunta: ‘En caso de una protesta pública, ¿las autoridades deberían o no hacer excepciones en la aplicación de la ley, si los que protestan son campesinos?’⁶, se contestó como sigue (véase Cuadro I.2-2):

CUADRO I.2-2
¿Las autoridades deberían o no hacer excepciones en la aplicación de la ley si los que protestan son campesinos?

Sí	1815
En parte	967
No	2014
Otra	24

Introducción a la estadística: Un enfoque dirigido a la sociología.

FUENTE: *La Naturaleza del Compromiso Cívico*, IFE, UDES, IISUNAM. Estudio levantado entre personas mayores de edad en Octubre de 2003.

La proporción de ‘Sí’ es el número de personas que contestaron esa respuesta, entre el número total de personas que participaron, esto es: $1815/(1815+967+2014+24) = 1815/4820 = 363/964$.

³ Cuando se trabaja con proporciones, siempre se intentará reducir la fracción a su mínima expresión.

⁴ El símbolo representa precisamente al porcentaje, y siempre que acompañe a un número, significa que se refiere a una proporción, considerando que 100 es el total, o número máximo.

⁵ Éste símbolo representa también a la razón, y se lee como: (el primer número) a (el segundo número). En nuestro ejemplo: 49 personas dijeron ‘muy estable’ a (o contra) 24 que dijeron ‘nada estable’, y si se reduce la fracción a números decimales se encuentra la cantidad de veces que cabe la primera opción en la segunda.

⁶ FUENTE: *La Naturaleza del Compromiso Cívico*, IFE, UDES, IISUNAM. Estudio levantado entre personas mayores de edad en Octubre de 2003.

El porcentaje de 'Sí', respecto al total, es: $100 \cdot (363/964) = 100 \cdot (0.376556) = 37.6556\%$.

La razón de 'Otra' a 'En parte', es: $24/967$.

Por otro lado, la razón de las personas que contestaron 'En parte' en la segunda pregunta, a las que contestaron 'Poco estable', en la primera pregunta, es de: $967/207 = 967:207$. Con esto último, se demuestra que es posible, sólo con las razones, comparar dos números referentes a cosas distintas⁷.

2.2 Estructura, formato y lectura de los datos⁸.

Se dispone de cuatro métodos de representación estadística. Los datos pueden: incorporarse en un párrafo del texto; ponerse en forma tabular; colocarse en un arreglo semi-tabular; expresarse gráficamente.

- Representación escrita. La ventaja de éste sistema radica en que el autor puede dirigir la atención hacia determinadas cifras, haciéndolas resaltar así, y también puede llamar la atención sobre las comparaciones de importancia. Por otro lado, la combinación de cifras y texto, no es un sistema muy eficaz, ya que es necesario leer, o por lo menos registrar, todo el párrafo antes de que se pueda comprender el significado de todo el conjunto de cifras. La mayoría de las personas no pueden comprender fácilmente los datos cuando se expresan de esta manera, y resulta difícil para el lector aislar las cifras individuales.
- Representación tabular. Éste método para exponer los datos estadísticos suele ser superior al uso del texto. Un cuadro con su título se debe explicar por sí mismo, aún cuando puede ir acompañado frecuentemente de un párrafo que facilite la interpretación o que llame la

⁷ En realidad, sería mejor ver sus proporciones (o porcentajes) respecto al total al que corresponde cada número, para, así, después verificar el porcentaje que representa cada número dentro de su grupo, sino, se puede caer en un error de interpretación. La idea del enunciado es ejemplificar la posibilidad de llevar a cabo la razón, sin embargo, como el lector puede notar, no tiene mucho sentido la misma, principalmente porque la cantidad total o universo del que provienen es distinta.

⁸ Para el desarrollo del presente capítulo, véase [2], págs. 60-82.

atención sobre las cifras importantes. Un cuadro, por lo general, es mucho más breve que el texto. Puesto que las cifras deben aparecer sin texto, la presentación es más concisa. Un cuadro deberá tener por lo menos cuatro partes esenciales: título, columna matriz (columna de la izquierda y su encabezado), encabezados de las columnas y cuerpo. También puede contener una nota preliminar y una o más notas al calce. Si las cifras del cuadro no son originales, se incluye también una nota de la fuente, acompañada a veces de una nota preliminar, pero generalmente debajo del cuadro y debajo de las notas al calce del mismo, si las hay.

- Representación semitabular. Cuando en un estudio se van a usar solamente unas cuantas cifras, puede dividirse el texto, enumerándose los datos, y después presentándose los números. Este método no se usa a menudo, pero es útil por cuanto se hacen resaltar las cifras, lo que no sucedería si estuvieran intercaladas en una o dos frases en el propio texto. De este modo, las cifras pueden a veces compararse con mayor facilidad.

- Representación gráfica. Los artificios gráficos son extremadamente útiles y muy eficaces para presentar de golpe una cantidad limitada de información. En el capítulo I.6, se dará a detalle una explicación sobre la representación gráfica.

Desde el punto de vista del uso, hay dos tipos de cuadros. En primer lugar hay cuadros generales o de referencia, que se usan como una fuente de información. Estos son frecuentemente muy extensos y abarcan muchas páginas. En segundo lugar, hay cuadros de texto o de resumen, que de ordinario son de tamaño relativamente reducido y que se planean de manera que se haga resaltar con la mayor intensidad posible, un dato, o varios datos estrechamente relacionados.

Ejemplo.

El siguiente es un ejemplo de representación en forma de texto:

“..., de las 455 quejas que se recibieron, 130 se encuentran en trámite en alguna de las etapas del procedimiento de investigación que marca la ley, se han concluido 96 expedientes que se encuentran pendientes de resolución o en los que ya existe un pronunciamiento por parte del Ombudsman. En 47 expedientes de queja se logró llegar a la resolución mediante la conciliación propuesta por la Comisión a los quejosos y autoridades denunciadas, apelando a la buena fe de las autoridades administrativas y dando con ello una pronta respuesta ante el reclamo de la ciudadanía. Por otra parte éste organismo ha recibido en lo que va del año 68 quejas correspondientes al ámbito de competencia de la Comisión Nacional de los Derechos Humanos, lo que ha permitido el pronto acceso a dicho órgano nacional por parte de los habitantes de Aguascalientes. Dentro del proceso de investigación esta Comisión se preocupa por fomentar la participación activa de los quejosos, en lo que va del año se archivaron 51 expedientes en los que se decretó la falta de interés por parte del peticionario cuando ante el llamado de la Comisión no se dio respuesta a diferentes requerimiento y comunicados, en 13 asuntos se determinó que no existía materia para que la Comisión se pudiera pronunciar, mientras que en 43 expedientes se presentó un desistimiento por parte de los quejosos.”⁹

El párrafo anterior también se puede presentar en forma tabular, y quedaría como sigue (véase Cuadro I.2-3):

CUADRO I.2-3
Quejas presentadas*

TIPO DE SEGUIMIENTO	NUMERO DE QUEJAS
Trámite	130
Investigación concluida	96
Conciliados	47
No ratificados	8
Incompetencia	46
Turnados a CNDH	22
Archivado por falta de interés	50
Desistimiento	43
Falta de materia	13
Total	455

Introducción a la estadística: Un enfoque dirigido a la sociología.

**FUENTE: Informe Anual del Presidente 2003. Comisión Estatal de los Derechos Humanos del Estado de Aguascalientes. Págs.8-9.*

⁹ FUENTE: Informe Anual del Presidente 2003. Comisión Estatal de los Derechos Humanos del Estado de Aguascalientes. Págs.- 8-9.

En este caso, tenemos el título (“*Quejas presentadas*”), la columna matriz (la columna “*tipo de seguimiento*”), el encabezado de todas las columnas (“*tipo de seguimiento*”, “*número de quejas*”), el cuerpo (que son los tipos y sus números respectivos). Por último, como los datos no son originales de éste texto, se ha indicado la fuente.

Por último, veamos ahora un ejemplo semitabular.

“El total de quejas presentadas ante la Comisión Estatal de los Derechos Humanos del Estado de Aguascalientes fue de 455. De éstas:

- 130 están en trámite.
- En 96 casos fue concluida la investigación.
- 47 fueron conciliados.

Con esto se puede observar que el 28.57% del total de quejas aún están en trámite, lo cual indica que...”¹⁰

Como se puede ver, no se usan todos los datos, sólo los de interés en cierto momento, además se utiliza mucho texto, y, aún así, los datos que resaltan son los números, ya que el resto del texto, para alguien que busca los datos, resulta irrelevante.

2.3 Construcción de indicadores demográficos y sociales.

La historia de los indicadores corre paralela a la historia de las ciencias sociales. Las ciencias sociales aparecen a finales del siglo XIX en un contexto positivista, que hizo que estas nuevas disciplinas nacieran con la necesidad de cuantificar sus objetos de estudio. Esta obligación constituía todo un reto, puesto que hasta entonces parecía que sólo era posible realizar medidas físicas, propias de los objetos de estudio de las ciencias naturales. La psicofísica se encargó de establecer un puente entre las disciplinas naturales y sociales permitiendo cuantificar las sensaciones a partir de un referente físico conocido.

¹⁰ FUENTE: Informe Anual del Presidente 2003. Comisión Estatal de los Derechos Humanos del Estado de Aguascalientes. Págs.- 8-9.

Como se ha manejado en el transcurso de éste capítulo, siempre será bueno resumir la información estadística que se recopile. Los indicadores son una de las maneras más eficientes de resumir los datos. Como ejemplo, si suponemos que en un país existen 1000 mujeres entre 15 y 49 años, y después de un censo se descubre que ellas han tenido, en total, 1500 hijos, entonces se dice que en ese país hay un índice de fecundidad general de 1.5. Pero para llegar a esos resultados se debe seguir un sistema que nos permita saber que lo que obtenemos es confiable, a la vez que comparable con otro tipo de datos (uno puede inventar una forma novedosa para calcular el índice general de fecundidad, pero si nadie más lo usa, ese dato sería aislado e incomparable).

2.3.1 Reglas generales para la construcción de indicadores.

Supongamos que se nos dice que en Jamaica la población menor de 18 años es de 1,007,000 habitantes, mientras que esa misma población pero refiriéndonos a Alemania, es de 14,286,000 habitantes¹¹. Con lo visto hasta ahora por el lector, podemos decir que la población joven de Alemania es 14.2 veces mayor que la de Jamaica, o bien, que la población joven de Jamaica representa al 7.05% de la población joven de Alemania. Sin dejar de ser esto cierto, ¿nos atreveríamos a decir que en Alemania es muy grande la población de jóvenes, o bien que es muy pequeña la población de jóvenes de Jamaica? Si así lo hiciéramos, caeríamos en un grave error. El lector tal vez se pregunte porqué, o bien, dirá que se está contradiciendo al resto del capítulo. Ahora, consideremos que en Alemania existen 82.3 millones de habitantes, mientras que en Jamaica, la población total es de 2.7 millones de habitantes¹², se puede ver que, la población joven de Jamaica representa al 37% de los habitantes, mientras que en Alemania representa sólo al 17% de los habitantes. La diferencia es inmensa, pero, lo más importante, ¿cómo obtener estadísticas confiables y comparables, que representen realmente a la dimensión de lo que se quiere estudiar?

¹¹ FUENTE: UNICEF, 2007.

¹² FUENTE: Buró de Población de Estados Unidos. Datos de 2007.

Hay una gran cantidad de indicadores de uso relativamente común en el análisis de la estructura y dinámica de la población que pueden considerarse componentes de la cultura general de los científicos sociales. En cursos en los que se tratan los temas demográficos, los indicadores demográficos abarcan una parte importante del tiempo. Aquí sólo se darán algunos lineamientos generales que son útiles tanto para los indicadores demográficos como para los sociales.

- Se debe saber cuáles son los datos con los que se va a trabajar, y éstos datos deben estar disponibles en todas nuestras fuentes. Por ejemplo, si vamos a trabajar con el número de inmigrantes en un estado, entre enero y febrero, si queremos hacer comparaciones con los demás estados, o bien dentro de cada municipio, deberemos contar con los mismos datos en cada uno de los lugares de interés.
- Cada vez que se haga una relación, se deberá pensar en qué nos indica dicha relación, cuales serán sus valores máximos y mínimos, y éstos qué representan, por último, debemos saber si esta información nos es útil o no.
- Cada vez que se construya una relación, para compararla, debemos utilizar exactamente el mismo tipo de dato, sino se caería en lo que se denomina sesgo estadístico, y puede conducir a conclusiones falsas por parte del investigador.

Como ejemplo, veamos al índice de masculinidad. Éste índice se define como el cociente entre el número de hombres sobre el número de mujeres. Normalmente se expresa por cien y su resultado debe interpretarse como la cantidad de hombres por cada 100 mujeres. Expresándolo en fórmula, queda como sigue:

$$(N_{\text{hombres}} / N_{\text{mujeres}}) * 100.$$

Siguiendo las recomendaciones arriba mencionadas, podemos ver que:

- Para cada espacio en el que deseemos obtener éste índice, será necesario conocer el número exacto de hombres y el de mujeres.

- Se puede ver que el número mínimo posible en ese índice es el de 0 y representa a una población carente de hombres, de lo que se deduce que entre más cercano de 0 esté el número, menor será el número de hombres por cada 100 mujeres. Por el contrario, teóricamente, no es posible obtener un máximo, ya que para una población pequeña de mujeres (tendiente a 0), el índice puede hacerse muy grande, lo que indica al final, que entre mayor sea el índice, menor será la cantidad de mujeres¹³.

- Imaginemos que tenemos la cifra del número de hombres en el estado de Nayarit de 1999, y la del número de mujeres en el mismo estado, pero de 2000. Dado que tenemos ambos datos, se nos puede ocurrir la construcción del índice de masculinidad. Pero, ¿estaría bien?, si suponemos que la población de Nayarit va en franco crecimiento, nos daría un dato que no es cierto, en este caso, la construcción del índice nos otorgaría una estimación del dato real, debido a que los datos no empatan al 100%. Si aún así decidimos obtener el índice, debemos estar concientes de que no es el número real, y que podemos caer en errores si no se sigue una interpretación adecuada.

2.4 BIBLIOGRAFÍA

[1] **Blalock**, Hubert M. *Estadística Social*. Fondo de Cultura Económica. México, 1966.

[2] **Croxton**, Frederick E. *Estadística General Aplicada*. Fondo de Cultura Económica, México, 1959.

[3] **Mangin, Jean – Pierre Lévy; Varela Malloy, Jesús; Abad González, Julio**. *Análisis Multivariable para las Ciencias Sociales*. Pearson Educación, España, 2004.

¹³ Nuevamente, hay que señalar que estos datos se deben tomar con mucho cuidado, porque en este caso el que el número del índice sea grande no implica que el número de hombres sea muy grande, sino que el de mujeres es muy pequeño.

3. CONJUNTOS Y NÚMEROS

3.1 Conjuntos y Subconjuntos

Es posible definir, con una axiomática, los conceptos “conjunto, elemento y pertenencia”. Sin embargo, para los fines del presente texto, es suficiente considerar como primitivos dichos conceptos y manejarlos en la forma intuitiva usual; es decir, los elementos pertenecen a conjuntos y un conjunto está formado por todos sus elementos. Así pues, dos conjuntos son iguales si y sólo si tiene los mismos elementos¹.

Como notación, se usará a las letras mayúsculas (A, B, C, \dots) para representar a los conjuntos y a las letras minúsculas (a, b, c, \dots) para representar a los elementos de los conjuntos. Para expresar a un conjunto por medio de sus elementos se hará por medio de llaves. Por ejemplo, si un conjunto A tiene por elementos a ‘ a ’, ‘ b ’, ‘ c ’, ‘ d ’, y ‘ e ’, entonces el conjunto se expresa como sigue:

$$A = \{a, b, c, d, e\}.$$

Hay que dejar en claro que, en realidad, un conjunto puede ser casi cualquier cosa, así, se pueden definir: el conjunto de todas las mujeres; el conjunto de pobres en un país, el conjunto de población mayor de 70 años, etc. No obstante, desde el punto de vista matemático, la definición de un conjunto se debe hacer a través una abstracción que impida cualquier tipo de ambigüedad, lo cual se puede hacer mostrando todos sus elementos, por medio de una fórmula con símbolos lógicos, o bien combinando las dos maneras anteriores². Por otro lado, no siempre se podrán distinguir a todos los elementos del conjunto, por ejemplo, el conjunto de mujeres en una casa puede ser de dos elementos, a saber:

¹ Véase Cárdenas, 1999.

² Véase Enderton, 1977.

CONJUNTO DE MUJERES EN CASA = {María, Carmen}³.

Sin embargo, en conjuntos con muchos elementos, se hace muy difícil, o hasta imposible, expresarlos por medio de todos sus elementos. Por ejemplo, el conjunto de todas las mujeres de México, para ser expresado por medio de sus elementos, tendría que incluirlas a todas, lo cual resulta casi imposible. Aún peor, si consideramos al conjunto de gotas de agua en un litro de agua, resulta imposible definirlo por medio de sus elementos, debido a que a cada gota se le puede dividir en varias gotas, que, a su vez, pueden ser subdivididas, y así sucesivamente. En resumen, si bien es posible definir un conjunto a través de la descripción de sus elementos, casi siempre será mejor hacerlo a partir de una abstracción.

Es importante destacar que en los conjuntos no importa el orden en el que aparecen sus elementos, lo importante es que siempre estén éstos presentes. Regresando a nuestros ejemplos:

$$A = \{a, b, c, d, e\} = \{a, c, b, d, e\} = \{c, e, a, d, b\} \neq \{a, b, c, d, f\}^4.$$

CONJUNTO DE MUJERES EN CASA = {María, Carmen} = {Carmen, María}.

3.1.1 El conjunto universo y el conjunto vacío.

Hay dos conjuntos particularmente interesantes, el conjunto universo y el conjunto vacío. El conjunto universo se define como el conjunto que contiene a todos los demás conjuntos⁵. En este texto se expresará por medio de la letra griega omega mayúscula (Ω). El conjunto vacío se define como el conjunto que

³ Todos los elementos de un conjunto deben poder distinguirse, así, si en este caso las dos mujeres se llamaran María, deberíamos distinguirla de otra forma, si no lo hiciéramos, diríamos que ambas son la misma persona. Aquí se puede observar la utilidad de las fórmulas con elementos lógicos que impidan la ambigüedad.

⁴ Nótese que el elemento 'f' no aparecía en el conjunto A, y que no puede reemplazar a otro elemento faltante.

⁵ En términos estrictos, el conjunto universo no es un conjunto (Enderton, 1977), ya que puede incluir demasiados conjuntos. Por ejemplo, si estamos hablando de números, el conjunto universo puede contener letras, colores, figuras geométricas, etcétera. Cuando en este texto se hable del conjunto universo se hará referencia al conjunto de todos los conjuntos de la abstracción de la que se esté hablando (números, letras, países, personas).

no tiene elementos. Se denota normalmente por el símbolo \emptyset y, por definición, pertenece a todos los conjuntos⁶.

Pertenencia. Se dice que un elemento pertenece a un conjunto si y solo si es uno de sus elementos y se denota por el símbolo \in y, si el elemento no pertenece a dicho conjunto, se denota por \notin . Veamos algunos ejemplos.

Sea A el conjunto definido anteriormente, $a \in A$, pero $f \notin A$, y, tal y como acabamos de definir, $\emptyset \in A$.

3.1.2 Subconjuntos.

Se dice que B es un subconjunto de A , si B es un conjunto, y además, todos los elementos de B son elementos de A , y en este caso, se denota como $B \subset A$.⁷

Si hay algún elemento en B que no es elemento de A , entonces B no es subconjunto de A , y esto se escribe como $B \not\subset A$. Veamos algunos ejemplos.

Sean:

$A = \{\text{Las mujeres que viven en México}\}.$

$B = \{\text{Las mujeres que viven en Colima}\}.$

$C = \{\text{Las mexicanas que viven en Colima}\}.$

$D = \{\text{Las mexicanas}\}.$

Entonces, las siguientes relaciones son ciertas:

$$A \supset B, A \supset C, A \not\subset D. \quad ^8$$

$$B \subset A, B \supset C, B \not\subset D.$$

$$C \subset A, C \subset B, C \not\subset D.$$

⁶ Es natural poder seguir expresando al conjunto aún después de 'quitarle' a todos sus elementos, así, aunque se le 'extraigan' a todos sus elementos, el conjunto aún tiene al elemento vacío.

⁷ Nótese que la definición nos indica que todo conjunto es subconjunto de sí mismo. A los subconjuntos de un conjunto que no son el mismo conjunto se les llama subconjuntos propios. A los subconjuntos que un conjunto que son el mismo conjunto se les llama subconjuntos impropios.

⁸ Nótese que el símbolo de pertenencia se puede utilizar en cualquier sentido, siempre que indique a la relación correctamente.

$$D \not\subset A, D \not\subset B, D \not\subset C, D \subset D.$$

3.2 Operaciones con conjuntos.

Así como en la aritmética de los números se pueden realizar diversas operaciones (suma, resta, producto, división), entre conjuntos también existen operaciones: unión, intersección, complemento y diferencia.

3.2.1 Unión.

Se define la unión de dos conjuntos como la operación con la cual se genera un tercer conjunto que contiene a todos los elementos de ambos conjuntos y se denota por el símbolo \cup . Sean A, B, C , conjuntos cualesquiera, entonces la unión tiene las siguientes propiedades:

1. $A \subset A \cup B, B \subset A \cup B.$
2. $A \cup B = B \cup A.$
3. $(A \cup B) \cup C = A \cup (B \cup C).$

Sean:

$$A = \{\text{Las personas de raza negra}\}.$$

$$B = \{\text{Las personas de raza latina}\}.$$

$$C = \{\text{Los mexicanos}\}.$$

$$D = \{\emptyset\}.$$

A partir de esos conjuntos definamos:

$$E = \{A \cup B\}. \text{ Entonces, } E = \{\text{Las personas de raza negra o latina}\}.$$

$F = \{A \cup B \cup C\}$. Entonces, $F = \{\text{Las personas de raza negra, o latina, o los mexicanos}\}$.

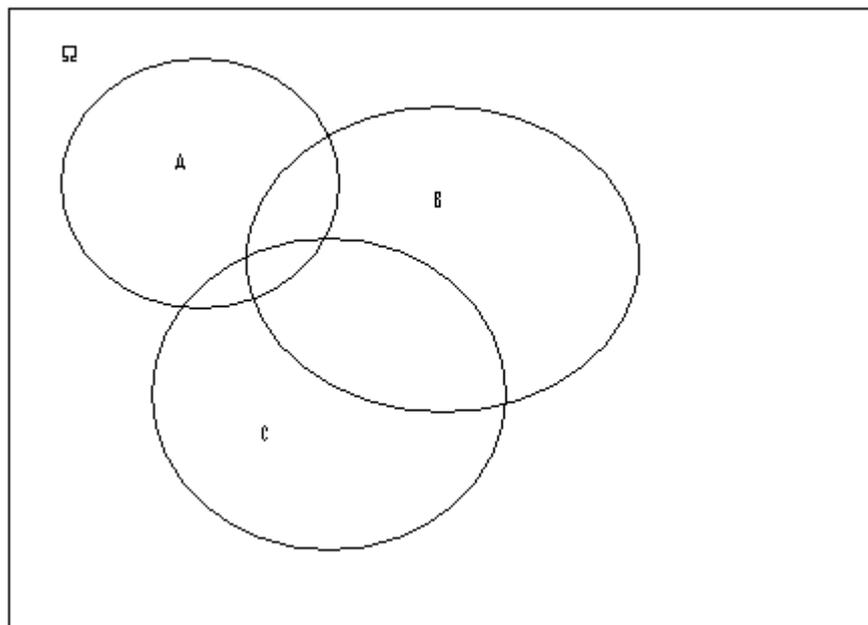
$G = \{A \cup B \cup D\}$. Entonces, $G = \{\text{Las personas de raza negra o latina}\}$. Además, es claro que $G = E$.

Por último, señalemos las siguientes relaciones:

$$A \subset E, B \subset E, A \subset F, B \subset F, C \subset F, A \subset G, B \subset G, D \subset G, D \subset E, D \subset F.$$

Como una ayuda gráfica para observar las operaciones entre conjuntos, es posible ‘verlos’ por medio de Diagramas de Venn, dibujando círculos que representen a los conjuntos e inscribiéndolos en el conjunto universo. Por ejemplo, con los conjuntos que ahora tenemos, en diagrama de Venn se observan como sigue:

Imagen I.3-1
Diagrama de Venn (1)

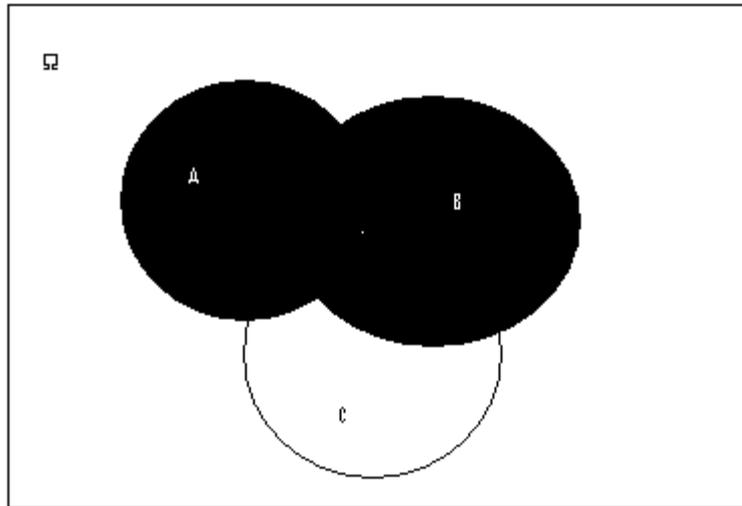


Introducción a la estadística: Un enfoque dirigido a la sociología.

Al conjunto E se le puede distinguir como sigue:

$$E = G =$$

Imagen I.3-2
Diagrama de Venn (2)

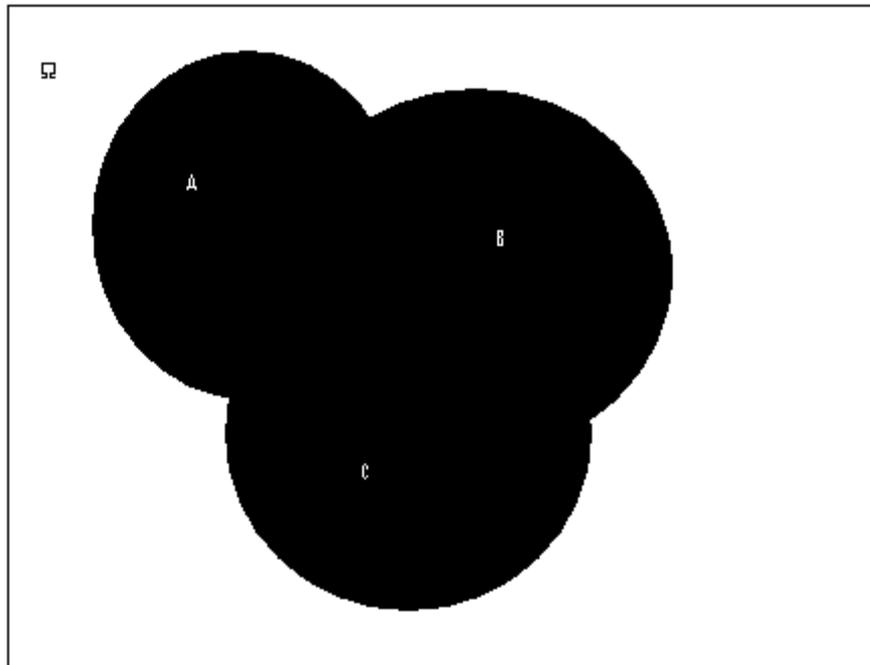


Introducción a la estadística: Un enfoque dirigido a la sociología.

Al conjunto F se le puede caracterizar como sigue:

$$F =$$

Imagen I.3-3
Diagrama de Venn (3)



Introducción a la estadística: Un enfoque dirigido a la sociología.

3.2.2 Intersección.

La intersección de dos conjuntos es el conjunto que contiene a los elementos que son parte del primer conjunto y parte del segundo conjunto y se denota por el símbolo \cap . Sean A, B, C , conjuntos cualesquiera, entonces la intersección tiene las siguientes propiedades:

1. $A \cap B \subset A, A \cap B \subset B$.
2. $A \cap B = B \cap A$.
3. $A \cap (B \cap C) = (A \cap B) \cap C$.

Sean los conjuntos:

$A = \{\text{Los diputados}\}$.

$B = \{\text{Los miembros del Congreso}\}$.

$C = \{\text{Los militantes del PRI}\}$.

$D = \{\text{Los militantes del PRD}\}$.

$E = \{\emptyset\}$.

Si definimos:

$F = \{A \cap B\}$. Entonces $F = \{\text{Los miembros del Congreso que son diputados}\}$.

$G = \{A \cap B \cap C\}$. Entonces $G = \{\text{Los miembros del Congreso que son diputados miembros del PRI}\}$.

$H = \{C \cap D\}$. Entonces $H = \{\text{Los militantes del PRI que son militantes del PRD}\}$.

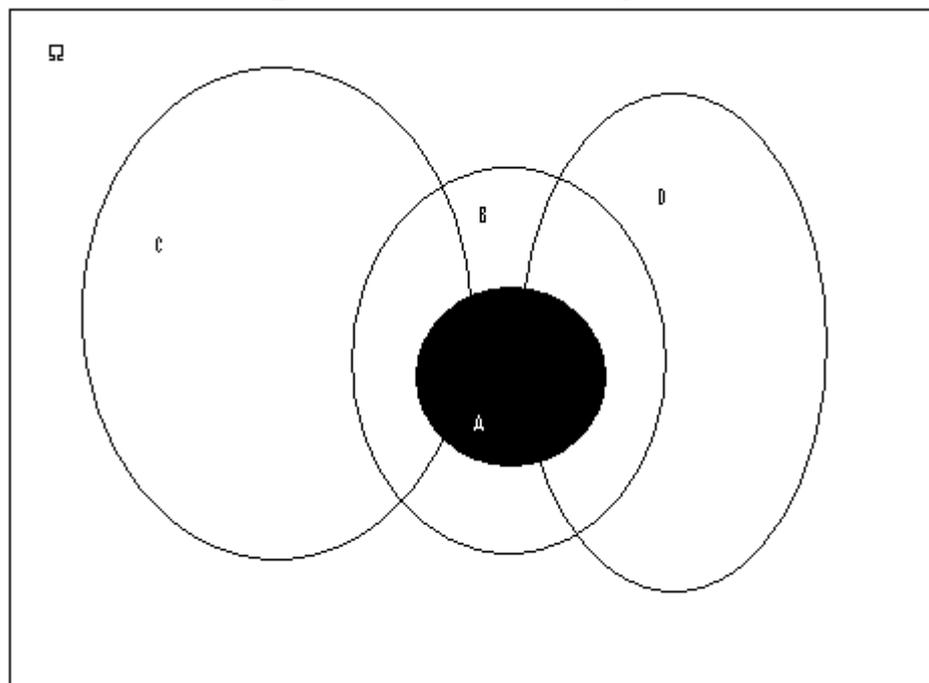
$I = \{C \cap E\}$. Entonces $I = \{\text{Los militantes del PRI que no están en el universo}\} = \emptyset$.

Por último, observemos las siguientes relaciones:

$$F \subset A, F \subset B, G \subset A, G \subset B, G \subset C, I \subset C.$$

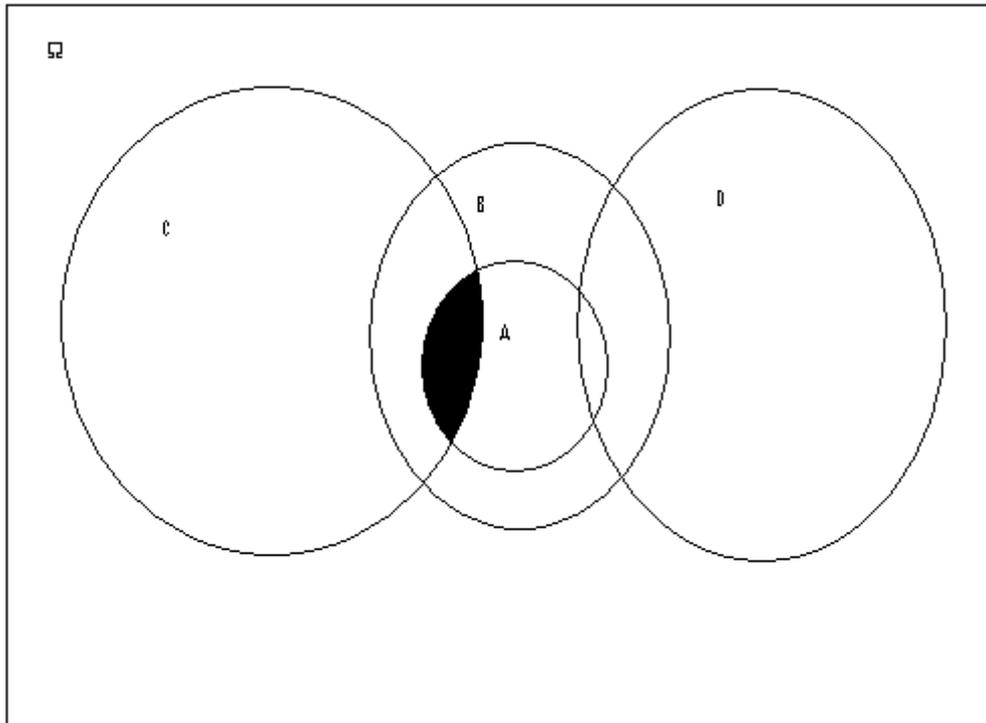
Si expresamos a los conjuntos por medio de diagramas de Venn, quedan como sigue.

Imagen I.3-4
Diagrama de Venn del conjunto F



Introducción a la estadística: Un enfoque dirigido a la sociología.

Imagen I.3-5
Diagrama de Venn del conjunto G



Introducción a la estadística: Un enfoque dirigido a la sociología.

3.2.3 El complemento.

Se define al complemento de un conjunto como el conjunto que contiene a todos los elementos que no están en el primero, pero sí en el conjunto universo (recordemos que el conjunto universo es aquel que contiene a la totalidad de lo que hablamos). Si A es un conjunto cualquiera, entonces, su complemento se denota como A^{C^9} .

Sean:

$$\Omega = \{\text{Los mexicanos}\}.$$

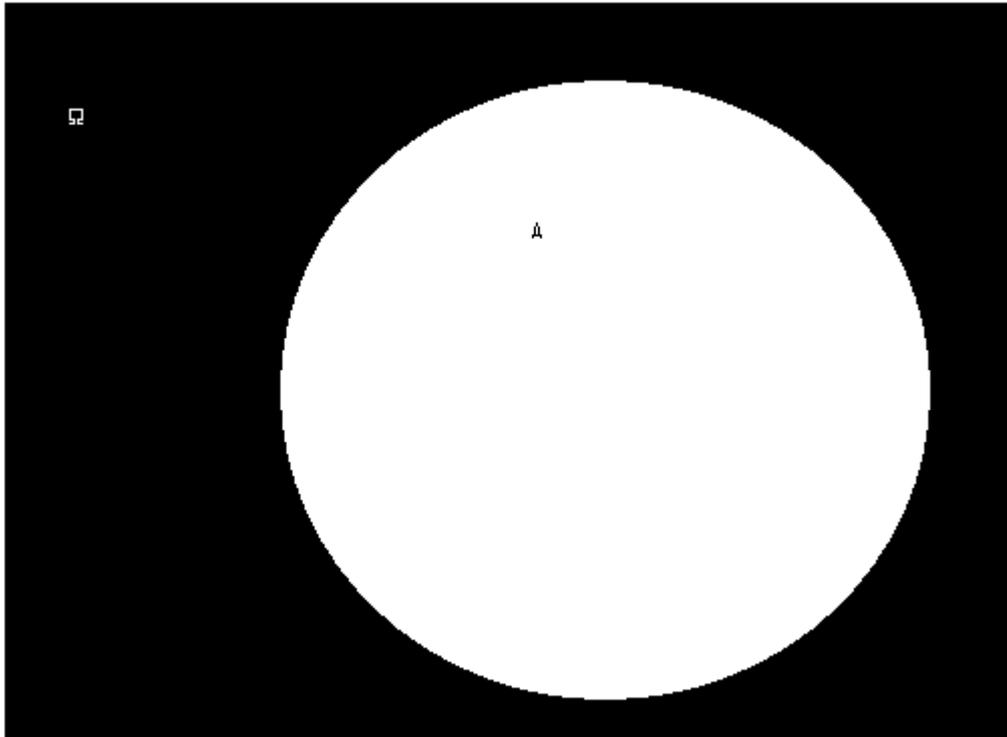
$$A = \{\text{Los mexicanos pobres}\}.$$

⁹ En algunos casos, también llega a ser denotado como A^c .

Entonces $A^C = \{\text{Los mexicanos que no son pobres}\}$.

Expresado por medio de diagramas de Venn:

Imagen I.3-6
Diagrama de Venn (4)



Introducción a la estadística: Un enfoque dirigido a la sociología.

Sean A , B , C tres conjuntos cualesquiera, las siguientes cuatro propiedades son muy importantes, ya que relacionan las operaciones vistas hasta ahora (unión, intersección y complemento). A las primeras dos se les conoce como leyes distributivas y a las dos últimas se les conoce como leyes de D'Morgan.

1. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
2. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
3. $(A \cup B)^C = A^C \cap B^C$.
4. $(A \cap B)^C = A^C \cup B^C$.

Si bien la demostración de éstas cuatro propiedades es relativamente sencilla, no es objeto del presente realizarla, pero los siguientes ejemplos las clarificarán.

Sean:

$A = \{\text{Los trabajadores de una fábrica que trabajan 8 horas}\}.$

$B = \{\text{Los trabajadores de una fábrica que están de vacaciones}\}.$

$C = \{\text{Los trabajadores de una fábrica que ganan más de \$6000 mensuales}\}.$

1. Por la propiedad 1, el conjunto resultante es el de los trabajadores que trabajan 8 horas y los que están de vacaciones o ganan más de \$6000 mensuales. Ese conjunto también puede ser expresado como el de los trabajadores que están de vacaciones y trabajan 8 horas, o los que ganan más de \$6000 mensuales y trabajan 8 horas, con lo que queda claro que la proposición es cierta.

2. Por la propiedad 2, el conjunto resultante es el de los trabajadores que trabajan 8 horas o los que están de vacaciones y ganan más de \$6000 mensuales. Ese conjunto también puede ser expresado como el de los trabajadores que trabajan 8 horas o están de vacaciones y los trabajadores que trabajan 8 horas o ganan más de \$6000 mensuales, con lo que queda claro que la proposición es cierta.

3. Por la propiedad 3, el conjunto resultante es el de los trabajadores en los que no es cierto que trabajen 8 horas o estén de vacaciones. Si no es cierto que trabajan 8 horas o están de vacaciones, entonces, ni trabajan 8 horas ni están de vacaciones (ya que si trabajaran 8 horas o estuvieran de vacaciones lo que no es cierto, sería cierto), con lo que queda claro que la proposición es cierta.

4. Por la propiedad 4, el conjunto resultante es el de los trabajadores en los que no es cierto que están de vacaciones y trabajan 8 horas. Si no es cierto que trabajan 8 horas y están de vacaciones, entonces, o no trabajan 8 horas o no están de vacaciones, con lo que queda claro que la proposición es cierta.

Con las operaciones entre conjuntos vistas hasta ahora, es suficiente para relacionar cualesquiera dos conjuntos, sin embargo, por facilidad, se define la diferencia.

3.2.4 Diferencia.

La diferencia entre dos conjuntos cualesquiera A y B se define como el 'quitarle' a A todos los elementos que también están en B , y se denota como $A - B$.

Se puede notar que $A - B = A \cap B^c$.

3.3 Una construcción conjuntista de los números.

Desde el inicio del pensamiento del hombre, ha sido necesario contar algunas cosas. Con la necesidad de contar se da inicio a lo que ahora conocemos como matemáticas. El progreso de las matemáticas ha sido muy largo, y lo que ahora conocemos como número no necesariamente es lo que se conoció como número en un principio. Un ejemplo claro es que los romanos no conocían al número cero (0), ni a los números negativos.

Cuando estamos en la enseñanza básica se nos enseñan primeramente a los números con los que podemos contar y establecer pertenencias, así, ejemplos como el tener 3 manzanas, 5 lápices o 10 pesos, entre otros, nos son mostrados como la importancia de conocer a los números. Éstos números son descritos en la actualidad como los números naturales, y a finales del siglo XIX se estableció una manera conjuntista de establecerlos. El conjunto de los número naturales se definió primariamente como:

$$\mathbf{N} = \{1, 2, 3, 4, 5, \dots\}.$$

Con el progreso de nuestros estudios, se vuelve natural el querer expresar la carencia de ciertos objetos, así, deseamos saber cómo expresar

que no tenemos manzanas, lápices o dinero. De ahí, la necesidad del número cero (0). Éste número es adicionado al conjunto de los números naturales, con lo que finalmente obtenemos la definición actual de los números naturales:

$$\mathbf{N} = \{0, 1, 2, 3, 4, 5, \dots\}^{10}.$$

Resulta natural querer asociar al conjunto vacío con el 0. De hecho, desde el punto de vista del conteo, si un conjunto no tiene elementos (el conjunto vacío), se dice que tiene 0 elementos, si posee un elemento, se dice que tiene 1 elemento, y así sucesivamente, y es por eso que al 0 se le considera como un número natural. En general, a los números naturales se les define como:

$$0 = \emptyset.$$

$$1 = \{0\}.$$

$$2 = \{1, 0\}.$$

$$3 = \{2, 1, 0\}.$$

$$4 = \{3, 2, 1, 0\}.$$

Y así sucesivamente.

Con el progreso de nuestros estudios, se vuelve natural el querer expresar la operación por medio de la cual nos quitan cosas. Así, si tenemos 3 manzanas y damos 2, nos queda 1 manzana, si tenemos 10 pesos y pagamos 8 por un artículo, nos quedan 2 pesos. Para poder expresar este tipo de operaciones se crea a la resta o diferencia. Pero llega el momento en que deseamos saber que pasa si tenemos 3 manzanas y necesitamos dar 4 manzanas, es decir, necesitamos dar más de lo que tenemos, o abstrayéndolo más, adquirimos una deuda. Los números naturales entonces ya no nos bastan, y necesitamos otro tipo de números, a los cuales llamamos números negativos. Al conjunto que resulta de la unión de los números naturales y los números negativos se le llama el conjunto de los números enteros, y se expresa de la siguiente manera:

¹⁰ En algunas las ramas de las matemáticas no se considera al 0 como un número natural, pero, en teoría de conjuntos, sí es considerado como tal.

$$\mathbf{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}.$$

Sin embargo llega el momento en que estos números no nos son suficientes. Nos enfrentamos a la necesidad de dar media manzana, o bien de tener 3 pesos con 50 centavos, es común que la gente pida un cuarto de kilo de jamón, medio kilo de huevo, etcétera. Dada esa necesidad humana, se crearon los ahora llamados números racionales, que conjuntivamente se pueden expresar como:

$$\mathbf{Q} = \frac{p}{q}, p \in \mathbf{Z}, q \in (\mathbf{N} - \{0\}).$$

Cuando hemos avanzado más en nuestros estudios, nos encontramos con que existen números que no pueden ser expresados ni siquiera en forma racional (aquí es importante destacar, para quien aún no lo ha notado, que todos los números enteros son racionales, es decir, que los números enteros son un subconjunto propio de los números racionales, en términos de conjuntos, $\mathbf{N} \subset \mathbf{Z} \subset \mathbf{Q}$), por ejemplo, la cantidad de veces que cabe el diámetro de un círculo en su perímetro, o la medida de la diagonal de un cuadrado de tamaño uno. Así, llegamos a números como $\pi, \sqrt{2}$, etc. Con ellos nacen los números irracionales¹¹, y se denotan por \mathbf{I} .

Con esos números se completa lo que se conoce como la recta real, que contiene a todos los números más utilizados¹². Así, los números reales se definen como:

$$\mathbf{R} = \mathbf{Q} \cup \mathbf{I}.$$

3.4 Igualdades y desigualdades¹³.

¹¹ Aunque parezca increíble, la cantidad de números irracionales es mucho mayor que la de números racionales.

¹² En la actualidad, existen otros números aún más complejos en su forma de expresión, éstos números tienen una gran aplicación en áreas como la física o las ingenierías, sin embargo, a lo largo del texto no serán usados, y además, se considera que el otorgar una definición de los mismos, puede confundir más de lo que ayuda a la comprensión del texto.

¹³ El desarrollo general de éste capítulo, se basa en Haaser 1999 págs. 34 – 38.

A continuación, se darán algunas propiedades sobre desigualdades. Así como por medio de la pertenencia se pueden comparar dos conjuntos, las desigualdades sirven para comparar dos números. Para ello se hará la siguiente afirmación:

Para toda pareja de números reales, a , b , se cumple una y sólo una de las siguientes propiedades¹⁴:

1. $a < b$, que se lee como “ a es menor que b ”.
2. $a = b$.
3. $b < a$.

Aparte del símbolo $<$ es conveniente emplear $>$, \geq y \leq . $a > b$, que se debe leer “ a es mayor que b ”, tiene el mismo significado que $b < a$; $a \geq b$ significa que “ a es mayor que b o a es igual a b ”; y $a \leq b$ significa “ a es menor que o igual a b ”.

Un número a se llama positivo si $a > 0$ y negativo si $a < 0$. No se debe creer que si hay un signo menos precediendo a un símbolo que representa a un número, como $-a$, por ello $-a$ ha de ser negativo. Por ejemplo, si $a = -3$, entonces $-a = 3$, un número positivo. Se dice que dos números tienen el mismo signo si ambos son positivos o ambos son negativos y que son de signo diferente si uno es positivo y el otro negativo. A continuación, se darán una serie de propiedades, las cuales no se demostrarán para evitar pérdida de claridad, pero se ejemplificará su uso.

1. Si $a < b$ y $c < d$, entonces $a + c < b + d$.
2. Si $a < b$, entonces $-a > -b$.
3. Si $a < b$ y $c < 0$, entonces $ac > bc$ (la expresión ac representa al producto de a por c , también denotado como axc o $a*c$).
4. Si $a \neq 0$, entonces $a^2 > 0$.
5. Si $0 \leq a < b$ y $0 \leq c < d$, entonces $ac < bd$ (la expresión $0 \leq a < b$ significa que $0 \leq a$ y $a < b$).

¹⁴ A este principio se le conoce como el principio de tricotomía y es uno de los 13 axiomas de los números reales.

6. Si a y b tienen el mismo signo, entonces $ab > 0$. Si a y b son de diferente signo, entonces $ab < 0$.
7. Si $a \neq 0$, $a^{-1} = 1/a$ tiene el mismo signo que a .
8. Si a y b tiene el mismo signo y $a > b$ entonces $a^{-1} < b^{-1}$.
9. Si $a \geq 0$ y $b \geq 0$, entonces $a^2 > b^2$ si $a > b$. Si $a > b$, entonces $a^2 > b^2$.
10. Si $b \geq 0$, entonces $a^2 > b$ si y sólo si $a > \sqrt{b}$ ó $a < -\sqrt{b}$.
11. Si $b > 0$, entonces $a^2 < b$ si y sólo si $-\sqrt{b} < a < \sqrt{b}$.

Ejemplos

1. Sean $a = 10$, $b = -20$, $c = 5$, $d = -3$. Es claro que $b < a$, y que $d < c$. Entonces, por la propiedad 1, $b + d < a + c$. $b + d = -20 + -3 = -23$. $a + c = 10 + 5 = 15$, y $-23 < 15$, tal y como la propiedad establecía.
2. Sean a y b como en el ejemplo anterior. Es claro que $b < a$, y siguiendo la propiedad 2, entonces $-a < -b$. $-a = -10$, $-b = -(-20) = 20$, y $-10 < 20$, con lo que se cumple la propiedad.
3. Sean a , b y d como en el ejemplo 1. $b < a$, y como $d < 0$, por la propiedad 3, $bd > ad$. $bd = -20 * -3 = 60$, $ad = 10 * -3 = -30$, y $60 > -30$, con lo que se cumple la propiedad.
4. Sean a , b , c y d como en el ejemplo 1. $a^2 = 10 * 10 = 100 > 0$, $b^2 = -20 * -20 = 400 > 0$, $c^2 = 5 * 5 = 25 > 0$, $d^2 = -3 * -3 = 9 > 0$. Como se puede observar, la propiedad cuatro se cumple.
5. Sean $a = 0$, $b = 3$, $c = 15$, $d = 16$. Es claro que $0 \leq a < b$ y $0 \leq c < d$, entonces, por la propiedad 5, $ac < bd$, $ac = 0 * 15 = 0$, $bd = 3 * 16 = 48$, y $0 < 48$, con lo que se cumple la propiedad 5.
6. Sean b y c como en el ejemplo anterior. Es claro que b y c tienen el mismo signo, entonces por la propiedad 6, $bc > 0$, $bc = 3 * 15 = 45 > 0$, con lo que se cumple la propiedad 6. También se puede apreciar que $-b$ y c tiene signos opuestos, por lo que $(-b)*c < 0$ por ésta propiedad, $(-b)*c = -3 * 15 = -45 < 0$.
7. Sean a y b como en el ejemplo 1. $a > 0$, y $a^{-1} = 1/a = 1/10 = 0.1 > 0$. $b < 0$, y $b^{-1} = 1/(-20) = -0.05 < 0$, con lo que se cumple la propiedad 7.

8. Sean c y d como en el ejemplo 5. Como c y d tiene el mismo signo y $c < d$, por la propiedad 8 $c^{-1} > d^{-1}$. $c^{-1} = 15^{-1} = 1/15 = 0.0667$, $d^{-1} = 16^{-1} = 1/16 = 0.0625$, y $0.0667 > 0.0625$, con lo que se cumple ésta propiedad.
9. Sean c y d como en el ejemplo 5. Como c y d son mayores que cero, y $c^2 = 15^2 = 225$, y $d^2 = 16^2 = 256$, es decir $225 = c^2 < d^2 = 256$, con lo que se cumple la propiedad 9.
10. Sean c y d como en el ejemplo 5. Es claro que $d > 0$, $c^2 = 225 > d$, y entonces, $15 > \sqrt{16}$ ó $15 < -\sqrt{16}$, con lo que se cumple la propiedad 10, al cumplirse la primera afirmación (hay que tener en cuenta que la conjunción ó nos indica algo parecido a una unión de conjuntos, es decir, con que pertenezca a uno de los dos, pertenece a la unión).
11. Sean b y d como en el ejemplo 5. Se puede verificar que $b^2 = 9 < 16 = d$. Entonces por la propiedad 11, $-\sqrt{16} = -4 < 3 < \sqrt{16} = 4$, con lo que se verifica la propiedad 11.

3.5 BIBLIOGRAFIA

[1] **Cárdenas**, Humberto, et al. *Álgebra Superior*. Trillas, 2ª edición. México, 1999.

[2] **Enderton**, Herbert. *Elements of Set Theory*. Academic Press inc, New York, 1977.

[3] **Haaser**, Norman B, et al. *Análisis Matemático*, volumen 1. Trillas, 2ª edición. México, 1999.

I.4. INDICES DE PRECIOS Y CANTIDAD

Al paso de los años los números índice han llegado a ser cada vez más importantes para la administración como indicadores de la cambiante actividad económica o de negocios. Los números índices, constituyen un sencillo artificio para comparar los términos de una o varias series cronológicas; considerando ésta última como una sucesión de observaciones de una variable tomada en instantes sucesivos.

Un número índice es una medida estadística diseñada para poner de relieve cambios en una variable o en un grupo de variables relacionadas con respecto al tiempo, situación geográfica, ingreso o cualquier otra característica. En el capítulo anterior se dio un panorama general acerca de los índices en demografía. A continuación veremos algunos de los índices más usados.

4.1 Índices de Precios.

4.1.1 Índice de Laspeyres

Es un índice de precios o producción que pondera a los precios respecto a un año original (o año base). El índice de producción puede ser expresado mediante la siguiente fórmula:

$$I_{Laspeyres} = \frac{\sum p_0 q_1}{\sum p_0 q_0},$$

Donde:

p_0 son los precios del año base de acuerdo al valor de las cantidades;

q_0 son las cantidades producidas en el año base; y,

q_1 son las cantidades producidas en el último año.

Veamos un ejemplo.

El siguiente tabulado (véase Cuadro I.4-1) muestra la producción de petróleo crudo en millones de barriles diarios, y el precio por barril.

**CUADRO I.4-1
Producción de petróleo tipo Maya en México**

AÑO	PRODUCCION*	DLS POR BARRIL**
1995	2617	14.42
1996	2858	17.37
1997	3022	14.89
1998	3070	8.72
1999	2906	14.52
2000	3012	22.87
2001	3127	17.31
2002	3177	20.93
2003	3371	24.22

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: INEGI, Información Estadística. Datos en miles de barriles al día.

**FUENTE: OPEC, Annual Statistical Bulletin, 2003 (Se refiere al petróleo mexicano tipo Maya).

En base a ésta información, si tomamos a 1995 como año base, el índice de Laspeyres para el año 1996 sobre la producción de petróleo mexicano será:

$$I_{1996} = \frac{(14.42dls)(2858*1000b / día)}{(14.42dls)(2617*1000b / día)} = 1.09209$$

Siguiendo la misma mecánica, el índice para cada uno de los años se presenta a continuación (véase Cuadro I.4-2):

**CUADRO I.4-2
Índice de Laspeyres para la producción de petróleo tipo Maya en México***

INDICE DE LASPEYRES PARA CADA AÑO, SOBRE EL PRECIO DEL PETRÓLEO							
1996	1997	1998	1999	2000	2001	2002	2003
1.092090	1.154757	1.173099	1.110432	1.150936	1.194880	1.213985	1.288116

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Se utilizan los mismos datos que en el Cuadro I.4-1.

El lector habrá notado que en éste ejemplo el índice únicamente representa al crecimiento (o decrecimiento) que tiene año con año (o periodo con periodo) la producción de petróleo. Sin embargo, el cálculo del índice se

vuelve más complejo (y también más útil) cuando se le añaden variables. Por ejemplo, pensemos que PEMEX¹ no se dedica únicamente a la venta de petróleo crudo, sino también a la venta de gasolinas, gas natural, gas licuado, derivados petroleros, etcétera. Un buen índice de Laspeyres que haga medición sobre la situación productiva de PEMEX habrá de considerar a toda su producción, y no solo a uno de sus productos. Veamos otro ejemplo.

El Índice Nacional de Precios al Consumidor (INPC) es un indicador que utiliza el Banco de México como referencia para calcular la inflación. En él se contemplan varios productos, a los cuales se les pondera para el cálculo del índice. Conforme aumentan de precio éstos productos, de acuerdo a la ponderación de los mismos, aumenta el índice. Dichos productos son agrupados por:

- Alimentos, bebidas y tabaco (22.74%).
- Ropa, calzado y accesorios (5.59%).
- Vivienda (26.41%).
- Muebles, aparatos y accesorios domésticos (4.86%).
- Salud y cuidado personal (8.58%).
- Transporte (13.41%).
- Educación y esparcimiento (11.54%).
- Otros servicios (6.88%)².

El porcentaje que aparece entre paréntesis representa la ponderación para cada rubro. Así, si para un periodo dado, aumentan 10% los precios del rubro de alimentos, bebidas y tabaco, en ese mismo periodo aumentará en un 2.274% el INPC ($22.74 * 0.10$), bajo el supuesto de que los demás precios permanecen sin cambio.

El siguiente cuadro (véase Cuadro I.4-3) muestra al INPC y al salario mínimo diario para la zona A de México, para el periodo comprendido entre 1995 y 1996.

¹ Petróleos Mexicanos, PEMEX, es la paraestatal mexicana que se dedica en forma exclusiva en el país a la producción de petróleo y de todos sus derivados y subproductos.

² FUENTE: Banco de México.

CUADRO I.4-3
Índice Nacional de Precios al Consumidos y Salarios Mínimos en el
Distrito Federal entre enero de 1995 y diciembre de 1996

Meses	INPC*	SM**
Ene / 1995	29.68	\$13.79
Feb / 1995	30.94	\$13.79
Mar / 1995	32.76	\$13.79
Abr / 1995	35.37	\$15.44
May / 1995	36.85	\$15.44
Jun / 1995	38.02	\$15.44
Jul / 1995	38.80	\$15.44
Ago / 1995	39.44	\$15.44
Sep / 1995	40.26	\$15.44
Oct / 1995	41.09	\$15.44
Nov / 1995	42.10	\$15.44
Dic / 1995	43.47	\$17.00
Ene / 1996	45.03	\$17.00
Feb / 1996	46.08	\$17.00
Mar / 1996	47.10	\$17.00
Abr / 1996	48.44	\$19.05
May / 1996	49.32	\$19.05
Jun / 1996	50.12	\$19.05
Jul / 1996	50.84	\$19.05
Ago / 1996	51.51	\$19.05
Sep / 1996	52.34	\$19.05
Oct / 1996	52.99	\$19.05
Nov / 1996	53.79	\$19.05
Dic / 1996	55.51	\$22.50

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Banco de México.

**FUENTE: Comisión Nacional de Salarios Mínimos.

De acuerdo al cuadro, entre enero de 1995 y diciembre de 1996, el INPC aumentó en un 87.03%, mientras que en el mismo periodo, el salario mínimo aumentó en un 63.32%.

Ahora, si consideramos al INPC para el caso de vivienda, al salario mínimo, y al índice de volumen de producción para la construcción, se tiene lo siguiente, de acuerdo al Cuadro I.4-4:

CUADRO I.4-4
INPC para viviendas, salarios mínimos e Índice de Volumen de Producción de febrero de 1995 a diciembre de 1996

Meses	INPCviv*	SM	IVPcons**
Feb / 1995	35.17	\$13.79	80.87
Mar / 1995	36.46	\$13.79	85.93
Abr / 1995	38.83	\$15.44	79.27
May / 1995	40.47	\$15.44	79.85
Jun / 1995	41.74	\$15.44	83.15
Jul / 1995	42.59	\$15.44	74.53
Ago / 1995	43.25	\$15.44	80.27
Sep / 1995	43.82	\$15.44	78.25
Oct / 1995	44.40	\$15.44	84.46
Nov / 1995	45.03	\$15.44	85.40
Dic / 1995	46.32	\$17.00	88.63
Ene / 1996	47.63	\$17.00	84.77
Feb / 1996	48.67	\$17.00	78.95
Mar / 1996	49.56	\$17.00	83.55
Abr / 1996	50.75	\$19.05	81.27
May / 1996	51.62	\$19.05	87.18
Jun / 1996	52.29	\$19.05	95.35
Jul / 1996	53.14	\$19.05	94.89
Ago / 1996	53.80	\$19.05	96.38
Sep / 1996	54.36	\$19.05	93.33
Oct / 1996	54.86	\$19.05	100.00
Nov / 1996	55.39	\$19.05	97.68
Dic / 1996	56.81	\$22.50	99.81

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Índice Nacional de Precios al Consumidor para el rubro de vivienda. FUENTE: Banco de México.

**Índice de Volumen de la Producción Industrial, para el rubro de la construcción. FUENTE: Banco de México.

Y si suponemos que:

p_0 son los precios del mes base de acuerdo al valor de las cantidades (que en este caso será el valor del INPCviv de febrero de 1995);

q_0 es el costo de producción, que en este caso tomaremos como: las cantidades producidas en el mes base (para el ejemplo será el índice de volumen de la producción industrial, para el rubro de la construcción, bajo el supuesto de que el IVPviv es un buen estimador de la producción de vivienda), multiplicadas por el salario mínimo; y,

q_1 son las cantidades producidas en el último mes (igualmente multiplicadas por el salario mínimo), que estarán representadas por el IVPviv A vigente en cada mes.

Entonces, el cálculo del índice de Laspeyres conforme a la fórmula general se calcula obteniendo los siguientes resultados (véase Cuadro I.4-5):

CUADRO I.4-5
Índice de Laspeyres para los datos del Cuadro I.4-4

Meses	INPCviv	SM	IVPcons	IL*
Feb / 1995	35.17	\$13.79	80.87	1.00
Mar / 1995	36.46	\$13.79	85.93	1.06
Abr / 1995	38.83	\$15.44	79.27	1.10
May / 1995	40.47	\$15.44	79.85	1.11
Jun / 1995	41.74	\$15.44	83.15	1.15
Jul / 1995	42.59	\$15.44	74.53	1.03
Ago / 1995	43.25	\$15.44	80.27	1.11
Sep / 1995	43.82	\$15.44	78.25	1.08
Oct / 1995	44.40	\$15.44	84.46	1.17
Nov / 1995	45.03	\$15.44	85.40	1.18
Dic / 1995	46.32	\$17.00	88.63	1.35
Ene / 1996	47.63	\$17.00	84.77	1.29
Feb / 1996	48.67	\$17.00	78.95	1.20
Mar / 1996	49.56	\$17.00	83.55	1.27
Abr / 1996	50.75	\$19.05	81.27	1.39
May / 1996	51.62	\$19.05	87.18	1.49
Jun / 1996	52.29	\$19.05	95.35	1.63
Jul / 1996	53.14	\$19.05	94.89	1.62
Ago / 1996	53.80	\$19.05	96.38	1.65
Sep / 1996	54.36	\$19.05	93.33	1.59
Oct / 1996	54.86	\$19.05	100.00	1.71
Nov / 1996	55.39	\$19.05	97.68	1.67
Dic / 1996	56.81	\$22.50	99.81	2.01

Introducción a la estadística: Un enfoque dirigido a la sociología.

**Índice de Laspeyres. Se usan los datos del Cuadro I.4-4, la construcción del índice es propia.*

Una interpretación del índice nos indicaría que los costos de producción de vivienda se duplicaron entre febrero de 1995 y diciembre de 1996.

4.1.2 El índice de Paasche

El índice de Paasche es un índice de producción o de precios que pondera a los precios respecto al año, o periodo manejado, en curso. Tiene la siguiente fórmula:

$$I_{Paasche} = \frac{\sum p_i q_i}{\sum p_0 q_i}$$

Donde,

p_0 son los precios o cantidades de producción del año o periodo base.

p_i son los precios o cantidades de producción del periodo i -ésimo.

q_i son los ponderadores del periodo i -ésimo, o bien, los precios de producción del periodo i .

Revisemos los ejemplos utilizados para el índice de Laspeyres.

Si utilizamos los datos del Cuadro I.4-1, y nuevamente tomamos a 1995 como año base, el índice de Paasche para 1996, será:

$$I_{1996} = \frac{(17.37 \text{usd} / \text{barril}) * (2858 \text{barriles})}{(14.42 \text{uds} / \text{barril}) * (2858 \text{barriles})} = 1.2046$$

Siguiendo la misma mecánica, el índice para cada uno de los años se presenta a continuación (véase Cuadro I.4-6):

**CUADRO I.4-6
INDICE DE PAASCHE PARA CADA AÑO, SOBRE EL PRECIO DEL
PETRÓLEO***

1995	1996	1997	1998	1999	2000	2001	2002	2003
1.0000	1.2046	1.0326	0.6047	1.0069	1.5860	1.2004	1.4515	1.6796

Introducción a la estadística: Un enfoque dirigido a la sociología.

**Índice de Paasche. Se usan los datos del Cuadro I.4-1, la construcción del índice es propia.*

El lector habrá notado que en éste ejemplo, el índice únicamente representa al crecimiento (o decrecimiento) que tiene año con año (o periodo con periodo) el precio del petróleo. Sin embargo, y al igual que con el índice de Laspeyres, el cálculo del índice se vuelve más complejo (y también más útil) cuando se le añaden variables. Veamos ahora el ejemplo referente al INPC para vivienda, haciendo las mismas consideraciones que para el índice de Laspeyres (véase Cuadro I.4-7):

p_0 son los precios del mes base de acuerdo al valor de las cantidades (que en este caso será el valor del INPCviv de febrero de 1995);

q_0 es el costo de producción, que en este caso tomaremos como: las cantidades producidas en el mes base, que en este caso será el índice de volumen de la producción industrial, para el rubro de la construcción (vamos a trabajar bajo el supuesto de que el IVPviv es un buen estimador de la producción de vivienda), multiplicadas por el salario mínimo; y,

q_1 son las cantidades producidas en el último mes (igualmente multiplicadas por el salario mínimo), que estarán representadas por el IVPviv A vigente en cada mes.

CUADRO I.4-7
Índice de Paasche para los datos del Cuadro I.4-4

Meses	INPCviv	SM	IVPcons	IP*
Feb / 1995	35.17	\$13.79	80.87	1.00
Mar / 1995	36.46	\$13.79	85.93	1.04
Abr / 1995	38.83	\$15.44	79.27	1.10
May / 1995	40.47	\$15.44	79.85	1.15
Jun / 1995	41.74	\$15.44	83.15	1.19
Jul / 1995	42.59	\$15.44	74.53	1.21
Ago / 1995	43.25	\$15.44	80.27	1.23
Sep / 1995	43.82	\$15.44	78.25	1.25
Oct / 1995	44.40	\$15.44	84.46	1.26
Nov / 1995	45.03	\$15.44	85.40	1.28
Dic / 1995	46.32	\$17.00	88.63	1.32
Ene / 1996	47.63	\$17.00	84.77	1.35
Feb / 1996	48.67	\$17.00	78.95	1.38
Mar / 1996	49.56	\$17.00	83.55	1.41
Abr / 1996	50.75	\$19.05	81.27	1.44
May / 1996	51.62	\$19.05	87.18	1.47
Jun / 1996	52.29	\$19.05	95.35	1.49
Jul / 1996	53.14	\$19.05	94.89	1.51
Ago / 1996	53.80	\$19.05	96.38	1.53
Sep / 1996	54.36	\$19.05	93.33	1.55
Oct / 1996	54.86	\$19.05	100.00	1.56
Nov / 1996	55.39	\$19.05	97.68	1.57
Dic / 1996	56.81	\$22.50	99.81	1.62

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Índice de Paasche. Se usan los datos del Cuadro I.4-4, la construcción del índice es propia.

De este tabulado se puede observar que el índice de Paasche aumenta en menor medida que el de Laspeyres. Pasemos ahora a un caso hipotético.

Supongamos que una familia necesita semanalmente para su subsistencia los siguientes bienes:

- Agua.
- Carne de pollo.
- Carne de res.
- Cebolla.
- Chile.
- Gas.
- Jitomate.

De acuerdo a sus necesidades, los precios de dichos artículos a lo largo de 10 semanas ha variado (véase Cuadro I.4-8). A partir de esos datos obtendremos los índices de Laspeyres y Paasche para la variación de precios.

CUADRO I.4-8
Variación de precios durante diez semanas para artículos seleccionados*

	PRECIOS						
	Agua (litro)	Carne de pollo (kg)	Carne de res (kg)	Cebolla (kg)	Chile (kg)	Gas (kg)	Jitomate (kg)
SEMANA 1	\$1.00	\$37.00	\$50.00	\$3.00	\$20.00	\$3.15	\$3.00
SEMANA 2	\$1.00	\$35.00	\$55.00	\$3.00	\$22.00	\$3.15	\$5.00
SEMANA 3	\$1.00	\$37.00	\$52.00	\$3.00	\$20.00	\$3.15	\$3.00
SEMANA 4	\$1.00	\$38.50	\$53.00	\$4.00	\$25.00	\$3.15	\$7.00
SEMANA 5	\$1.00	\$38.50	\$55.00	\$4.00	\$22.00	\$3.15	\$5.00
SEMANA 6	\$1.00	\$38.50	\$54.50	\$4.00	\$23.00	\$3.15	\$10.00
SEMANA 7	\$1.25	\$37.00	\$53.00	\$4.00	\$25.00	\$3.15	\$8.50
SEMANA 8	\$1.25	\$37.00	\$55.00	\$4.50	\$20.00	\$3.15	\$4.00
SEMANA 9	\$1.25	\$39.00	\$52.00	\$3.50	\$20.00	\$3.25	\$12.00
SEMANA 10	\$1.25	\$38.00	\$55.00	\$4.00	\$21.50	\$3.25	\$2.50

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Caso hipotético.

Asimismo, las necesidades de consumo de la familia, han variado de semana en semana como lo muestra el Cuadro I.4-9.

CUADRO I.4-9
Variación en el consumo de una familia durante diez semanas para artículos seleccionados*

	CANTIDADES							Cantidades
	Agua (litro)	Carne de pollo (kg)	Carne de res (kg)	Cebolla (kg)	Chile (kg)	Gas (kg)	Jitomate (kg)	
SEMANA 1	200.00	1.00	2.00	1.00	0.50	10.00	3.00	217.50
SEMANA 2	215.00	0.75	2.25	0.50	0.10	11.00	2.00	231.60
SEMANA 3	205.00	1.50	2.00	0.50	0.10	13.00	2.50	224.60
SEMANA 4	210.00	1.25	1.75	1.00	0.00	10.00	1.50	225.50
SEMANA 5	200.00	0.50	1.50	0.00	0.20	9.00	2.00	213.20
SEMANA 6	195.00	2.00	2.00	1.00	0.10	5.00	1.00	206.10
SEMANA 7	190.00	1.50	2.50	1.25	0.30	10.00	1.00	206.55
SEMANA 8	192.00	0.75	2.75	0.50	0.10	11.00	1.50	208.60
SEMANA 9	195.00	1.00	2.00	0.50	0.50	13.00	0.75	212.75
SEMANA 10	201.00	1.00	2.00	1.00	0.40	11.00	2.00	218.40

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Caso hipotético.

Por la variación en los precios, dicha familia ha tenido que gastar diferentes sumas en esos artículos (véase Cuadro I.4-10), misma que ha variado entre \$344.50 y \$463.90.

CUADRO I.4-10
Variación en el gasto de una familia durante diez semanas para artículos seleccionados*

	GASTO	Diferencia respecto a la semana anterior	Diferencia respecto a la semana 1
SEMANA 1	\$390.50		
SEMANA 2	\$413.35	5.85%	5.85%
SEMANA 3	\$416.45	0.75%	6.65%
SEMANA 4	\$396.88	-4.70%	1.63%
SEMANA 5	\$344.50	-13.20%	-11.78%
SEMANA 6	\$413.05	19.90%	5.77%
SEMANA 7	\$478.00	15.72%	22.41%
SEMANA 8	\$463.90	-2.95%	18.80%
SEMANA 9	\$449.75	-3.05%	15.17%
SEMANA 10	\$452.60	0.63%	15.90%

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Caso hipotético. Se utilizan los datos presentados en los Cuadros I.4-8 y I.4-9.

Las fórmulas para calcular los índices de Laspeyres y Paasche para la semana 2, tomando como base a la semana 1, son las siguientes:

$$I_{Laspeyres} = \frac{\left((200lt / agua)(\$1.00) + (1kg / pollo)(\$35.00) + (2kg / res)(\$55.00) + (1kg / cebolla)(\$3.00) \right. \\ \left. + (0.5kg / chile)(\$22.00) + (10kg / gas)(\$3.15) + (3kg / jitomate)(\$5.00) \right)}{\left((200lt / agua)(\$1.00) + (1kg / pollo)(\$37.00) + (2kg / res)(\$50.00) + (1kg / cebolla)(\$3.00) \right. \\ \left. + (0.5kg / chile)(\$20.00) + (10kg / gas)(\$3.15) + (3kg / jitomate)(\$3.00) \right)}$$

$$I_{Paasche} = \frac{\left((215lt / agua)(\$1.00) + (0.75kg / pollo)(\$35.00) + (2.25kg / res)(\$55.00) \right. \\ \left. + (0.50kg / cebolla)(\$3.00) + (0.10kg / chile)(\$22.00) + (11kg / gas)(\$3.15) \right. \\ \left. + (2kg / jitomate)(\$5.00) \right)}{\left((215lt / agua)(\$1.00) + (0.75kg / pollo)(\$37.00) + (2.25kg / res)(\$50.00) \right. \\ \left. + (0.50kg / cebolla)(\$3.00) + (0.10kg / chile)(\$20.00) + (11kg / gas)(\$3.15) \right. \\ \left. + (2kg / jitomate)(\$3.00) \right)}$$

Siguiendo esas fórmulas para todas las semanas, los valores que toman los índices son los siguientes (véase Cuadro I.4-11):

CUADRO I.4-11

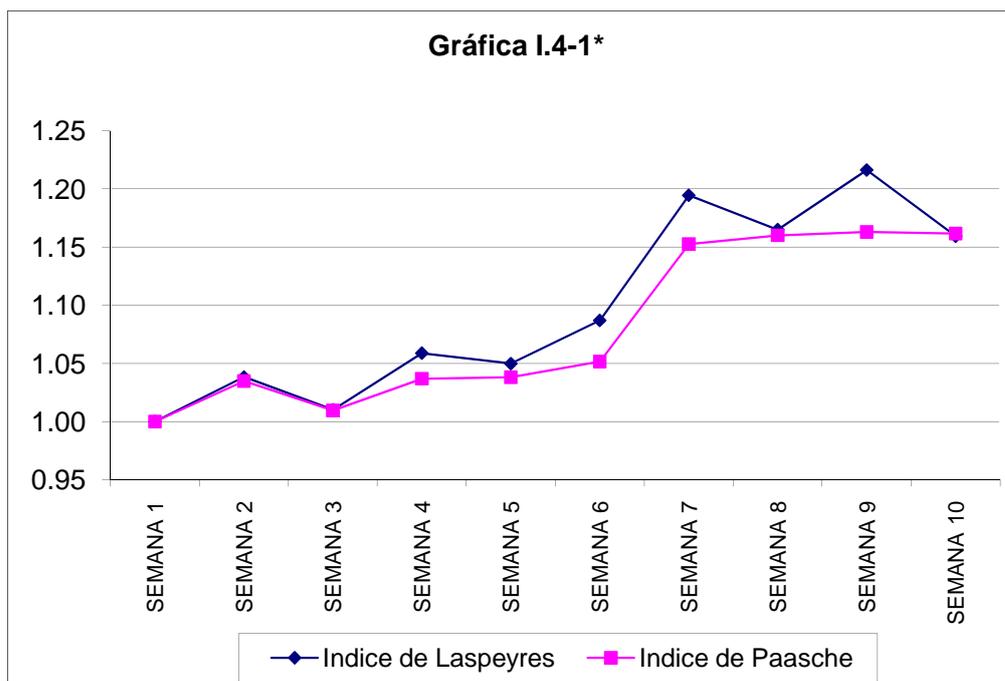
Índices de Laspeyres y Paasche para la economía de una familia*

Semana	Índice de Laspeyres	Índice de Paasche
SEMANA 1	1.0000	1.0000
SEMANA 2	1.0384	1.0349
SEMANA 3	1.0102	1.0097
SEMANA 4	1.0589	1.0369
SEMANA 5	1.0499	1.0381
SEMANA 6	1.0871	1.0517
SEMANA 7	1.1946	1.1525
SEMANA 8	1.1652	1.1600
SEMANA 9	1.2164	1.1630
SEMANA 10	1.1594	1.1616

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Caso hipotético. Se utilizan los datos presentados en los Cuadros I.4-8, I.4-9 y I.4-10.

Como se puede observar, el mayor aumento respecto a la semana 1, se da en la semana 7, cuando los precios en términos generales aumentaron 22.41%, mientras que la disminución más importante ocurrió en la semana 5, con una disminución de 11.78%. Por otro lado, el índice de Laspeyres alcanza su máximo valor en la semana 9, con un valor de 1.2165, al igual que el índice de Paasche, éste último, con un valor de 1.1630. Es fácil observar que el índice de Paasche es menor al de Laspeyres, excepto en la semana 10. Si los Graficamos, los valores de los índices se ven según la Gráfica I.4-1.



Introducción a la estadística: Un enfoque dirigido a la sociología.
 *Caso hipotético. Se utilizan los datos presentados en el Cuadros I.4-11.

En el gráfico se pueden observar varias cosas:

- Ambos índices dan un salto importante entre la semana 6 y la semana 7, que es cuando se da el mayor aumento respecto a la semana 1, que es nuestra semana base, es decir, ambos índices son sensibles al aumento de precios y cantidades.
- En la semana 3 se observa una disminución de ambos índices respecto a la semana 2, esto se debió a que tanto los precios en términos generales, como las cantidades, disminuyeron en dicha semana.
- En la semana 5 se presenta la mayor disminución de precios respecto a la semana 1, lo cual se nota en el índice de Laspeyres, mientras que el índice de Paasche aumenta. Esto significa que mientras los precios de la semana 5 considerando las cantidades de la semana 1 fueron inferiores a los precios de la semana 4 considerando las cantidades de la semana 1, mientras que los precios de la semana 5 considerando las cantidades de la misma semana, respecto a los precios de la semana base considerando las cantidades de la semana 5

fueron superiores a los precios de la semana 4 considerando las cantidades de esta semana (la 4), respecto a los precios de la semana base considerando las cantidades de la semana 4. Esto es, el índice de Laspeyres bajó porque, al mantener las cantidades constantes, los precios entre la semana 4 y la semana 5 disminuyeron, y el índice de Paasche subió porque, al tener como denominador constante a los precios de la semana base, pero considerando a las cantidades de la semana 5, en términos generales, los precios ponderados por las cantidades subieron.

Como se puede ver, el índice de Paasche tiene una interpretación más compleja, razón por la cual es más usado el índice de Laspeyres. Por otro lado, el índice de Laspeyres se deja llevar en gran medida por los movimientos de los precios, sin importar que las cantidades hayan aumentado o disminuido, una cuestión que sí es considerada por el índice de Paasche.

4.1.3 El índice de Fisher

Es la media geométrica entre el índice de Laspeyres y el índice de Paasche. Por lo tanto, su fórmula es la siguiente:

$$I_{Fisher} = \sqrt{(I_{Laspeyres})(I_{Paasche})}.$$

Se puede considerar que los índices de Laspeyres y Paasche representan las cotas del comportamiento de los precios o producciones. Así, tal y como se vio, El índice de Laspeyres se afectado sustancialmente ante una variación de precios, mientras que el índice de Paasche se ve afectado en menor medida por la variación de precios, ya que para cada periodo las cantidades también varían, y, en términos generales, varían de forma inversamente proporcional a los precios (es decir, cuando los precios aumentan, las cantidades tienden a bajar, y mientras los precios bajan, las cantidades tienden a subir), pero varía sustancialmente cuando hay un gran movimiento en las cantidades. Es entonces necesario obtener un punto intermedio. Una primera aproximación sería un promedio simple, sin embargo,

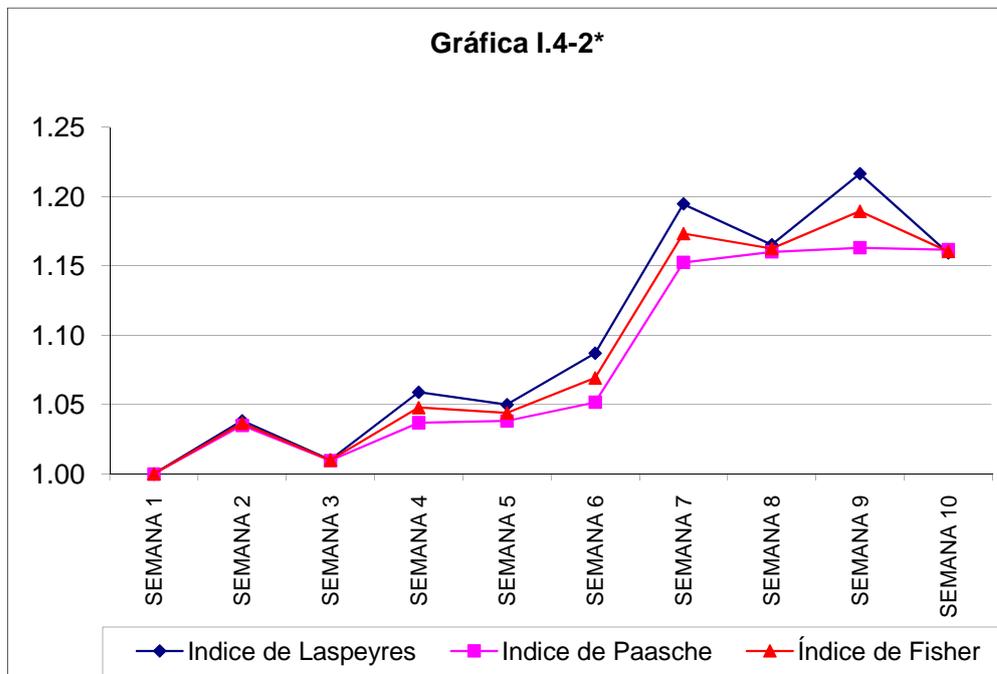
éste último se puede ver afectado por movimientos bruscos en uno de los dos índices (por ejemplo, si el índice de Laspeyres permanece constante, pero el de Paasche aumenta bruscamente, entonces el promedio simple también aumentará bruscamente). Es por ello que se prefiere la media geométrica, que tiende a suavizar los comportamientos bruscos. De nuestro ejemplo hipotético, el índice de Fisher sería (véase Cuadro I.4-12):

CUADRO I.4-12
Índices de Laspeyres, Paasche y Fisher para la economía de una familia*

	Índice de Laspeyres	Índice de Paasche	Índice de Fisher
SEMANA 1	1.0000	1.0000	1.0000
SEMANA 2	1.0384	1.0349	1.0367
SEMANA 3	1.0102	1.0097	1.0100
SEMANA 4	1.0589	1.0369	1.0478
SEMANA 5	1.0499	1.0381	1.0440
SEMANA 6	1.0871	1.0517	1.0692
SEMANA 7	1.1946	1.1525	1.1734
SEMANA 8	1.1652	1.1600	1.1626
SEMANA 9	1.2164	1.1630	1.1894
SEMANA 10	1.1594	1.1616	1.1605

Introducción a la estadística: Un enfoque dirigido a la sociología.
*Caso hipotético. Se utilizan los datos presentados en el I.4-12.

Y su gráfica es la siguiente (véase Gráfica I.4-2):



Introducción a la estadística: Un enfoque dirigido a la sociología.
*Caso hipotético. Se utilizan los datos presentados en el Cuadro I.4-12.

4.2 Inflación, deflación, poder adquisitivo, valor del peso.

A continuación se presentan algunos conceptos de gran importancia, tanto para la economía, como para la vida diaria.

4.2.1 Inflación

Es un aumento general sostenido en el nivel de precios que reduce el poder de compra que se tenía. Sus principales causas son: Aumento en la cantidad de dinero, exceso de demanda, aumento en los precios al público, la situación en el mercado laboral y cambios en los costos³.

En términos más sencillos, la inflación es el efecto por medio del cual ahora podemos comprar menos con el mismo dinero que antes. Es decir, si antes con \$100 comprábamos 100 manzanas, ahora, con \$100 compramos 96 manzanas. Sus causas también se pueden explicar de forma más sencilla.

- Cuando aumenta la cantidad de dinero significa que en términos generales, todos tenemos una mayor cantidad de dinero. Es decir, si en la colonia en la que vivimos, de repente a todos nos duplican el sueldo, vamos a poder comprar el doble de artículos en un principio, pero con el paso del tiempo, cuando los vendedores se den cuenta de que poseemos más dinero, ellos también van a aumentar sus precios, para poder obtener un mayor margen de ganancias, con lo que nuevamente perderemos poder adquisitivo, es decir, habrá inflación.

- Cuando hay exceso en la demanda significa que en términos generales, todos tenemos una mayor necesidad de los mismos productos. Por ejemplo, antes del inicio de clases, el número de personas que compra útiles escolares aumenta significativamente. Si suponemos que el número de cuadernos que queremos comprar es de 1000, pero sólo hay 500, entonces, el vendedor subirá los precios de los cuadernos para que sólo las personas con mayor poder de compra

³Cárdenas, 2000.

puedan llevarse los 500 cuadernos que tiene. Al igual que en el caso anterior, esto conlleva a una pérdida en el poder adquisitivo, con la consecuente inflación.

- La situación laboral también es importante, ya que, cuando existe una tasa de desempleo alta, los salarios tienden a bajar. En estas circunstancias, la capacidad de compra de la gente disminuye, y con eso, los artículos tienden a no aumentar de precio, para poder mantener los niveles de ventas, ya que es preferible vender 1000 artículos con un margen de ganancia del 15% a vender 100 artículos con un margen de ganancia del 50%. Por ejemplo, si cada artículo cuesta \$100, se vendería a \$115 en el primer caso, y a \$200 en el segundo caso, dejando ganancias de \$15000 en el primer caso y de \$10000 en el segundo caso.

- Por último, un cambio en los costos también puede generar inflación. Siguiendo el ejemplo anterior, si ahora el costo de producir 1000 artículos no es de \$100, sino que es de \$110, para mantener el margen de ganancias del 15%, es necesario vender los artículos en \$126.50, teniendo como consecuencia que si los salarios se han mantenido constantes, el poder de compra disminuirá, y la inflación subirá.

4.2.2 Deflación

Cuando existe una disminución general, sostenida en el nivel de precios, que aumenta el poder de compra que se tenía inicialmente, se le llama deflación. Dicha disminución deberá mantenerse al menos en 2 periodos consecutivos para ser considerada deflación. En caso de que exista una disminución general en un periodo, y al siguiente haya un aumento general de precios, se dice que hubo una inflación negativa, mas no una deflación⁴.

⁴ Cárdenas, 2000.

4.2.3 Poder adquisitivo

Es la capacidad de compra que tiene una persona o un grupo como cantidad de bienes que en distintos momentos pueden comprarse con una suma de dinero constante⁵. Es decir, el poder adquisitivo es el que nos dice cuánto podemos comprar con una misma cantidad de dinero. Cuando hay inflación, el poder adquisitivo disminuye, mientras que cuando hay deflación, el poder adquisitivo aumenta.

4.2.4 Valor del peso

El valor de la moneda depende de varios factores. Antes de tratarlos, supongamos que existen dos tribus, la tribu A y la tribu B. Para que exista comercio entre ambas tribus es necesario que exista un acuerdo respecto al medio que se va a utilizar para intercambiar sus bienes. Supongamos que la tribu A tiene 100 lingotes de oro, mientras que la tribu B tiene 500 lingotes de oro (cada lingote del mismo valor). Los lingotes de oro, representan el respaldo económico de cada tribu. Así, si la tribu A emite 1000 monedas para su intercambio, cada moneda tiene un valor de un décimo de lingote de oro, mientras que si la tribu B emite 1000 monedas para su intercambio, cada moneda tiene un valor de un medio de lingote de oro. Por lo anterior, cada moneda de la tribu B equivale a 5 monedas de la tribu A, y si una persona de la tribu A desea comprar un artículo de la tribu B que tiene un valor de 4 monedas de dicha tribu, ésta persona necesitará pagar 20 monedas de su tribu para poder comprar el artículo deseado. Pues bien, ese es el valor de la moneda de la tribu A. El proceso explicado se seguía anteriormente para calcular el valor de las monedas, cada una de las monedas tenía que estar respaldada en oro. Actualmente, en lugar de lingotes de oro, el valor de una moneda está basado en las reservas de un país, depositadas en el Banco Mundial. Además de dichos depósitos, los gobiernos de las diferentes naciones pueden decidir el valor de su moneda mediante algún decreto, es decir, siguiendo con nuestro caso, que el gobierno de la tribu A puede decidir que por cada moneda de la

⁵ Ibid.

tribu B pagará dos monedas de la tribu A, en lugar de las 5 monedas que equivaldrían al valor real. Por otro lado, actualmente también se llega a determinar dicho valor de acuerdo a las leyes de la oferta y la demanda. Lo anterior quiere decir que los valores varían de acuerdo a las leyes del mercado. Por ejemplo, en diciembre la gente realiza una gran cantidad de compras que requieren de una compra de artículos de otras naciones. Así, generalmente, en diciembre el valor del dólar aumenta respecto al valor de peso. Por ejemplo, tenemos 1000 dólares y necesitamos convertirlos en pesos, deberemos comprar pesos, con lo cual aumentará el valor del peso, debido a que hay una mayor demanda de pesos.

4.3 Cambio de base; empalme de índices.

4.3.1 Cambio de Base

Como se vio en la sección 4.1, los índices en economía requieren de un valor inicial o año base. Con el transcurso de tiempo, puede ser necesario cambiar al periodo base utilizado para la construcción de un número índice. Las razones que pueden llevarnos a esa decisión varían dependiendo de la situación. La pregunta es, cómo realizar el cambio de base. Pues bien, si se debe mantener la misma lógica de los índices (es decir, que aún represente al índice de Laspeyres cuando desde el principio representaba al índice de Laspeyres, por ejemplo), es necesario recalcular las fórmulas, únicamente cambiando nuestro periodo base.

4.3.2 Empalme de índices

En ocasiones es necesario comparar dos índices. Sin embargo, las fórmulas para calcular los índices son distintas en muchos de los casos, y aún siendo las mismas, es probable que el conjunto de datos que utilicen sea distinto, o bien, que utilicen bases diferentes. Por ello es necesario encontrar una manera fácil de comparación, que a la vez realmente represente las diferencias que queremos hacer notar. A continuación se describe de forma ilustrativa una manera de estandarizar índices.

Para ello, fijémonos en el índice de Laspeyres calculado para el ejemplo hipotético de la sección 4.1 con los datos del Cuadro I.4-11. Además, supongamos que queremos cambiar la semana base, y queremos que la nueva semana base sea la semana 5, es decir, ahora queremos que el valor del índice para la semana 5 sea 1. Para ello, debemos calcular el porcentaje que representan las diferentes semanas respecto a la semana 5. Por ejemplo, el valor de la semana 1 es el 95.24% del valor de la semana 5 (véase Cuadro I.4-13). Para el cálculo de dichos porcentajes, la fórmula seguida es:

$$P = \left(\frac{x_i}{x_0} \right) * 100\% , \text{ donde}$$

P es el porcentaje.

x_i es el valor de la semana i -ésima.

x_0 es el valor de la nueva semana base.

CUADRO I.4-13
Índices de Laspeyres y diferencia respecto a la semana 5*

	Indice de Laspeyres	Diferencia porcentual respecto a la semana 5
SEMANA 1	1.0000	95.24%
SEMANA 2	1.0384	98.90%
SEMANA 3	1.0102	96.22%
SEMANA 4	1.0589	100.85%
SEMANA 5	1.0499	100.00%
SEMANA 6	1.0871	103.54%
SEMANA 7	1.1946	113.78%
SEMANA 8	1.1652	110.98%
SEMANA 9	1.2164	115.85%
SEMANA 10	1.1594	110.43%

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Caso hipotético. Se utilizan los datos presentados en el I.4-11.

Ahora notemos que el valor de la semana 5 es el 100% del valor de la semana base, que es la misma semana. Además, se puede observar que el valor de la semana 6 es el 103.54% del valor de la semana 5, por lo que, si la semana 5 tiene ahora el valor de 1, el valor de la semana 6 será ahora un 103.54% del valor de la semana 5, que es 1, por lo que el valor de la semana 6 será: $(1.0000) * (1.0354) = 1.0354$. Generalizando, el valor del índice en cada semana será:

$$X_i = \left(\frac{x_i}{x_0}\right) * X_0 = \left(\frac{x_i}{x_0}\right)(1) = \frac{x_i}{x_0}, \text{ donde}$$

X_i es el valor para la i -ésima considerando la nueva base,

X_0 es el valor de la nueva semana base, considerándose a sí misma, por lo que su valor es 1,

x_i es el valor de la semana i -ésima.

x_0 es el valor de la nueva semana base.

Así, los nuevos valores del índice se presentan en el Cuadro I.4-14. Si bien este no es el índice de Laspeyres, nos será útil para el empalme de los índices.

CUADRO I.4-14
Índices de Laspeyres con base en la semana 5*

	Índice de Laspeyres con base en la semana 1	Diferencia porcentual respecto a la semana 5	Índice con base en la semana 5
SEMANA 1	1.0000	95.24%	0.9524
SEMANA 2	1.0384	98.90%	0.9890
SEMANA 3	1.0102	96.22%	0.9622
SEMANA 4	1.0589	100.85%	1.0085
SEMANA 5	1.0499	100.00%	1.0000
SEMANA 6	1.0871	103.54%	1.0354
SEMANA 7	1.1946	113.78%	1.1378
SEMANA 8	1.1652	110.98%	1.1098
SEMANA 9	1.2164	115.85%	1.1585
SEMANA 10	1.1594	110.43%	1.1043

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Caso hipotético. Se utilizan los datos presentados en el I.4-11.

Ahora, si notamos que las fórmulas son generales (es decir, no dependen del ejemplo), las podemos aplicar al índice de Paasche, que se muestra en el Cuadro I.4-15.

CUADRO I.4-15
Índices de Paasche con base en la semana 5*

	Índice de Paasche con base en la semana 1	Diferencia porcentual respecto a la semana 5	Índice con base en la semana 5
SEMANA 1	1.0000	96.33%	0.9633
SEMANA 2	1.0349	99.69%	0.9969
SEMANA 3	1.0097	97.26%	0.9726
SEMANA 4	1.0369	99.88%	0.9988
SEMANA 5	1.0381	100.00%	1.0000
SEMANA 6	1.0517	101.31%	1.0131
SEMANA 7	1.1525	111.02%	1.1102
SEMANA 8	1.1600	111.74%	1.1174
SEMANA 9	1.1630	112.03%	1.1203
SEMANA 10	1.1616	111.89%	1.1189

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Caso hipotético. Se utilizan los datos presentados en el I.4-12.

Una vez que se tienen los valores de ambos índices, estamos en posición de empalmarlos, es decir, de poderlos comparar adecuadamente, lo cual se presenta en el Cuadro I.4-16.

CUADRO I.4-16
Empalme de los índices de Laspeyres y Paasche con base en la semana 5*

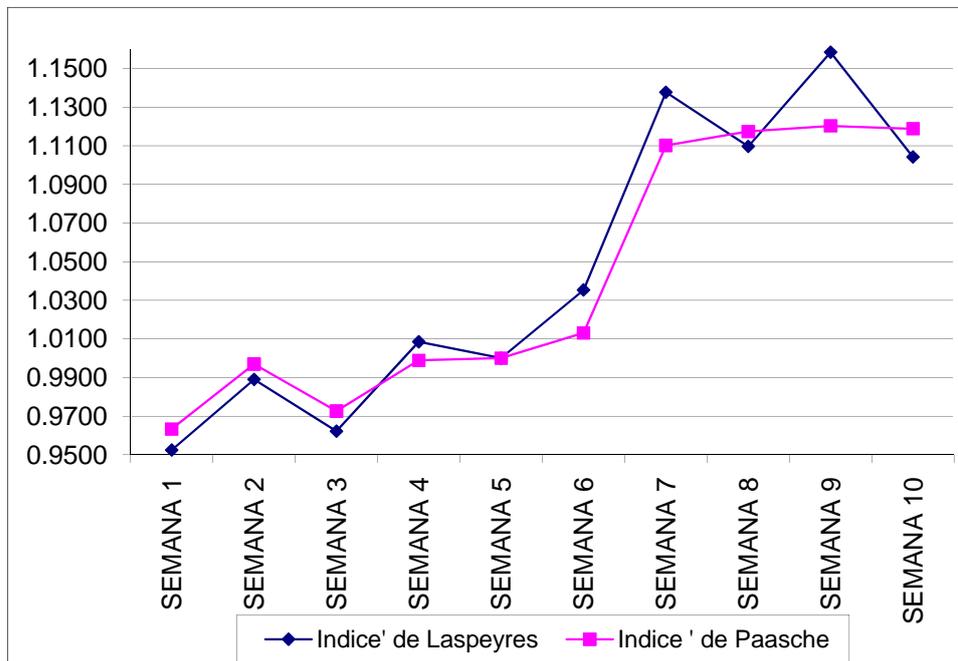
	Índice de Laspeyres	Índice de Paasche
SEMANA 1	0.9524	0.9633
SEMANA 2	0.9890	0.9969
SEMANA 3	0.9622	0.9726
SEMANA 4	1.0085	0.9988
SEMANA 5	1.0000	1.0000
SEMANA 6	1.0354	1.0131
SEMANA 7	1.1378	1.1102
SEMANA 8	1.1098	1.1174
SEMANA 9	1.1585	1.1203
SEMANA 10	1.1043	1.1189

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Caso hipotético. Se utilizan los datos presentados en el I.4-12.

Del cuadro debemos recordar que aunque se utiliza la palabra índice, no se están utilizando los índices conforme a las fórmulas ya vistas, sino que se está realizando una comparación entre ambos para una fácil comparación. Si los vemos gráficamente (véase Gráfico I.4-3), se puede observar que la gráfica cambia sustancialmente de forma, y se debe a que las comparaciones pierden la interpretación inicial de ambos índices. En cambio, ahora podemos observar que el índice de Paasche aumenta más en las semanas 1, 2, 3, 8 y 10 que el índice de Laspeyres, mismo que aumenta más que el de Paasche en las semanas 4, 6, 7 y 9.

GRÁFICO I.4-3*
Empalme de los índices de Laspeyres y Paasche*



Introducción a la estadística: Un enfoque dirigido a la sociología.
*Caso hipotético. Se utilizan los datos presentados en el I.4-16.

4.4 BIBLIOGRAFÍA

- [1] **Cárdenas** Cotiño, Gustavo A. *Diccionario de Ciencias Económico-Administrativas*. CUCEA, Universidad de Guadalajara, 2ª edición. México, 2000.
- [2] **Fox**, Karl A. *Manual de Econometría*. Amorrorto editores. Argentina, 1973.
- [3] **McAllister**, Harry E. *Elementos de estadística en la economía y los negocios*. ECASA, México, 1987.
- [4] **Salas**, Javier. *Econometría Aplicada a los países en desarrollo. El caso mexicano*. FCE, México, 1990.
- [5] **Gordon**, Robert J. *Macroeconomía*. Compañía Editorial Continental SA de CV. México, 1996.
- [6] **Scharles**, Maurice, et. al. *Economic Análisis, Theory and Applications*. Irwin inc, 6a edición. Estados Unidos, 1992.

I.5. Estadística Descriptiva.

5.1 Introducción.

En este capítulo se verán los conceptos principales de la estadística descriptiva. Su nombre se deriva de que éstos instrumentos nos permitirán resumir y describir una gran cantidad de datos, que de otra manera, sería muy difícil describir.

5.2 Tabulación de los datos.

La tabulación de los datos se refiere a la forma en que se acomodan los datos obtenidos. Siempre se buscará que su acomodo obedezca a reglas claras, y a un orden riguroso, que permitan al investigador observar con mayor facilidad la distribución general de los datos. Mientras que en la sección I.2-2 se analizaron las partes de un tabulado y el orden de presentación, aquí se profundizará acerca de la manera en que los datos son presentados para una lectura más ágil. Veamos algunos ejemplos.

5.2.1 Ejemplo 1.

Supongamos que en un domicilio se tienen diferentes tipos de mascotas, a saber, 1 perro, 2 gatos, 15 peces, 3 canarios y 2 tortugas. Entonces se pueden tabular sin ningún orden (véase Cuadro I.5-1), o bien, se pueden ordenar los datos de mayor a menor (véase Cuadro I.5-2), según el tipo de mascota, o bien, de menor a mayor (véase Cuadro I.5-3), o bien, por importancia relativa al hogar (véase Cuadro I.5-4).

CUADRO I.5-1
Mascotas en un hogar*

Tipo de Mascota	Cantidad
Perro	1
Gato	2
Pez	15
Canario	3
Tortuga	2

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Caso hipotético.

CUADRO I.5-2
Mascotas en un hogar*

Tipo de Mascota	Cantidad
Pez	15
Canario	3
Gato	2
Tortuga	2
Perro	1

Introducción a la estadística: Un enfoque dirigido a la sociología.
*Caso hipotético.

CUADRO I.5-3
Mascotas en un hogar*

Tipo de Mascota	Cantidad
Perro	1
Gato	2
Tortuga	2
Canario	3
Pez	15

Introducción a la estadística: Un enfoque dirigido a la sociología.
*Caso hipotético.

CUADRO I.5-4
Mascotas en un hogar*

Tipo de Mascota	Cantidad
Perro	1
Gato	2
Canario	3
Tortuga	2
Pez	15

Introducción a la estadística: Un enfoque dirigido a la sociología.
*Caso hipotético.

Si bien los cuatro cuadros contienen exactamente la misma información, es más fácil observar que en ese hogar existe una cantidad de peces superior a cualquier otra mascota si se observa el Cuadro I.5-2. Igualmente, si se observa el Cuadro I.5-3 es más fácil observar que en ese hogar sólo existe una mascota del tipo perro. Al observar el Cuadro I.5-4 resulta evidente, en cambio, observar que en esa casa el perro es la mascota que tiene mayor importancia en el hogar, en lugar de los peces, que tienen una importancia, relativa a ese hogar, muy pequeña. Si bien todas esas observaciones también se pueden realizar observando el Cuadro I.5-1 (excepto por la importancia relativa), si únicamente se tuviera a disposición esa tabla, resultaría más difícil llegar a dichas conclusiones.

5.2.2 Ejemplo 2.

El estado de Baja California Sur, México, tiene 5 municipios. Si se manejan únicamente los datos del nombre del municipio, y su población total, dichos datos pueden ser acomodados de diferentes maneras. En este ejemplo, se acomodarán alfabéticamente (véase Cuadro I.5-5), y de mayor a menor población (véase Cuadro I.5-6).

CUADRO I.5-5
Población de Baja California Sur por municipio*

Municipio	Población Total
Comondú	63,864
La Paz	196,907
Loreto	11,812
Los Cabos	105,469
Mulegé	45,989

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000¹.

CUADRO I.5-6
Población de Baja California Sur por municipio*

Municipio	Población Total
La Paz	196,907
Los Cabos	105,469
Comondú	63,864
Mulegé	45,989
Loreto	11,812

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000¹.

En el Cuadro I.5-5 se puede observar fácilmente que los municipios de Baja California Sur empiezan con la letra c, con la l, o con la m. En cambio, del Cuadro I.5-6, resulta fácil observar que el municipio más poblado es el de La Paz, y el menos poblado es el de Loreto. Seguramente al lector le parece más importante la información que resalta en el Cuadro I.5-6 a la que resalta del cuadro inmediato anterior. Es por ello que debemos ser cuidadosos respecto al orden en el que se tabulan los datos.

¹ El SCINCE es el Sistema de Consulta de Información Censal, que se basa en el XII Censo General de Población y Vivienda, del INEGI.

5.2.3 Ejemplo 3.

Hasta ahora, se han visto ejemplos donde tenemos dos columnas. Ahora veremos un ejemplo donde tenemos tres columnas. Ahora vamos a considerar a las 16 delegaciones del Distrito Federal de México y se considerará por delegación a la población ocupada que recibe menos de un salario mínimo mensual de ingreso por trabajo y a las viviendas particulares con techos de materiales ligeros, naturales y precarios. Se tabularan estos datos ordenando de mayor a menor la población ocupada que recibe menos de un salario mínimo mensual de ingreso por trabajo (véase Cuadro I.5-7), y luego se ordenarán de mayor a menor número de viviendas con techos de materiales ligeros, naturales y precarios (véase Cuadro I.5-8).

CUADRO I.5-7
Población que recibe menos de un salario mínimo y viviendas con materiales precarios por delegación en el Distrito Federal¹

Delegación	Pobl<1SM*	Viv c/mat prec**
Iztapalapa	71,799	64,833
Gustavo A. Madero	44,904	34,570
Álvaro Obregón	20,874	21,203
Tlalpan	19,918	24,838
Cuauhtémoc	19,104	4,194
Coyoacán	18,498	12,159
Venustiano Carranza	16,544	8,728
Xochimilco	15,567	20,008
Iztacalco	14,613	13,358
Azcapotzalco	13,342	13,172
Tláhuac	11,044	11,905
Miguel Hidalgo	8,985	8,127
Benito Juárez	8,891	2,384
La Magdalena Contreras	8,102	12,571
Milpa Alta	5,475	6,254
Cuajimalpa de Morelos	4,015	7,241

Introducción a la estadística: Un enfoque dirigido a la sociología.

¹FUENTE: SCINCE, 2000.

*Población ocupada que recibe menos de un salario mínimo mensual de ingreso por trabajo.

**Viviendas particulares con techos de materiales ligeros, naturales y precarios.

CUADRO I.5-8

Población que recibe menos de un salario mínimo y viviendas con materiales precarios por delegación en el Distrito Federal¹

Delegación	Pobl<1SM*	Viv c/mat prec**
Iztapalapa	71,799	64,833
Gustavo A. Madero	44,904	34,570
Tlalpan	19,918	24,838
Álvaro Obregón	20,874	21,203
Xochimilco	15,567	20,008
Iztacalco	14,613	13,358
Azcapotzalco	13,342	13,172
La Magdalena Contreras	8,102	12,571
Coyoacán	18,498	12,159
Tláhuac	11,044	11,905
Venustiano Carranza	16,544	8,728
Miguel Hidalgo	8,985	8,127
Cuajimalpa de Morelos	4,015	7,241
Milpa Alta	5,475	6,254
Cuauhtémoc	19,104	4,194
Benito Juárez	8,891	2,384

Introducción a la estadística: Un enfoque dirigido a la sociología.

¹FUENTE: SCINCE, 2000.

*Población ocupada que recibe menos de un salario mínimo mensual de ingreso por trabajo.

**Viviendas particulares con techos de materiales ligeros, naturales y precarios.

En el Cuadro I.5-7 resalta que las delegaciones Iztapalapa y Gustavo A. Madero son las que tienen un mayor número de población ocupada que recibe menos de un salario mínimo mensual de ingreso por trabajo, mientras que las delegaciones Milpa Alta y Cuajimalpa de Morelos son las que tienen un menor número de población con esas características. En cambio, en el Cuadro I.5-8 se puede observar que las delegaciones Iztapalapa y Gustavo A. Madero son también las que tienen un mayor número de viviendas particulares con techos de materiales ligeros, naturales y precarios, mientras que las delegaciones Cuauhtémoc y Benito Juárez son las que tienen un menor número de viviendas con esas características. En este caso resulta aún más clara la importancia que tiene el ordenamiento de las tablas para destacar datos de interés para la investigación.

5.3 Medidas de Posición.

Se les llama medidas de posición a las herramientas estadísticas que nos indican el comportamiento de los datos, respecto a algún parámetro. Estas

herramientas tienen una gran cantidad de aplicaciones. Una de las aplicaciones más importantes de las medidas de posición, se refiere al ingreso. En diferentes medios se dice que el algún porcentaje de la población gana menos de un dólar, en términos más generales, las medidas de posición nos pueden indicar cuánto gana el 10% mas pobre de una población, etcétera. Si bien este tipo de ejemplos son un poco más complejos y se tratarán en el capítulo posterior, las medidas de posición tienen una gran importancia. Antes de pasar a ellas, en las secciones 6.3.1 y 6.3.2 se dan las definiciones de datos sin agrupar y datos agrupados respectivamente; estas definiciones serán fundamentales en el resto del capítulo.

5.3.1 Datos sin agrupar.

Se consideran datos sin agrupar a los datos que hacen referencia a un valor específico. Por ejemplo, en la tabla 8 los datos no están agrupados (de hecho, todos los ejemplos de la sección 6.2 se refieren a datos sin agrupar), ya que cada valor representa a una delegación. Siempre es preferible para cálculos estadísticos tener a los datos sin agrupar, ya que con ellos se pueden realizar cálculos exactos.

5.3.2 Datos agrupados.

Se consideran datos agrupados a los datos que hacen referencia a varios valores. Por ejemplo, si quisiéramos agrupar al número de delegaciones con viviendas particulares con techos de materiales ligeros, naturales y precarios, lo podríamos hacer distinguiendo a las delegaciones con menos de 10,000 viviendas con esa característica, luego a las delegaciones con entre 10,000 y 20,000 viviendas con esa característica, luego a las delegaciones con entre 20,000 y 30,000 viviendas con esa característica, y por último, a las delegaciones con más de 30,000 viviendas con esa característica (véase Cuadro I.5-9). Si bien los cálculos realizados con datos agrupados suelen ser aproximaciones de datos reales, en muchos casos es necesario trabajar con ellos, ya sea porque no disponemos de los datos sin agrupar, o bien porque la

cantidad de datos sin agrupar es demasiado grande como para trabajar con ella.

CUADRO I.5-9
Delegaciones con techos de materiales ligeros en el Distrito Federal*

Intervalo	Número de Delegaciones*
Menos de 10,000 viviendas**	6
De 10,000 a 19,999 vivienda**s	5
De 20,000 a 29,999 viviendas**	3
30,000 viviendas o más**	2

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000.

**Se refiere a viviendas particulares con techos de materiales ligeros, naturales y precarios.

5.3.3 Los percentiles

Los percentiles son 99 valores de la variable estadística X que dividen a la distribución de frecuencias en 100 partes iguales². Es decir, cuando tenemos una serie de datos, los percentiles nos dividen a esa serie en 100 partes iguales. El cálculo de los percentiles resulta importante para la división de los datos. Por ejemplo, el percentil 60 nos dice que debajo de él se encuentra el 60% de las observaciones, y por encima de él el 40% de las observaciones. Algunos de los percentiles más importantes son el 25, el 50, y el 75. C_k determinará al percentil de orden k , es decir, que debajo del percentil k se encuentran $k\%$ de las observaciones. El cálculo de los percentiles para datos agrupados es el siguiente.

1. Determinar el número de observaciones que representan el $k\%$ de la distribución de frecuencias.
2. Calcular las frecuencias acumuladas, comenzando por el intervalo más pequeño. El percentil k , C_k , se encontrará en aquel intervalo cuya frecuencia acumulada supere el $k\%$ de las observaciones. En el caso de que la frecuencia acumulada coincidiera exactamente con el $k\%$ de las observaciones, C_k sería el límite superior del intervalo.
3. Una vez encontrado el intervalo crítico, aplicar la siguiente fórmula:

$$C_k = L_i + \frac{1}{n_i} \left(k * \frac{n}{100} - n_a \right), \text{ donde,}$$

² Véase Garrido, 1995.

L_i es el límite exacto inferior del intervalo crítico.

l es la amplitud real del intervalo.

n_i es la frecuencia del intervalo crítico.

n_a es la frecuencia acumulada hasta el intervalo anterior al intervalo crítico.

n es el número total de observaciones.

El cálculo de los percentiles para datos sin agrupar es el siguiente:

1. Ordenar los datos de menor a mayor.
2. Calcular la frecuencia acumulada de cada dato.
3. El percentil k , C_k se encontrará en el dato que tenga una frecuencia acumulada del $k\%$. En caso de que ningún dato cumpla con esa característica, se tomara al valor medio del dato que tenga una frecuencia acumulada menor a $k\%$, pero que el siguiente dato tenga una frecuencia acumulada mayor a $k\%$ y el dato que tenga una frecuencia acumulada mayor a $k\%$ pero que el dato anterior tenga una frecuencia acumulada menor a $k\%$.

Veremos un ejemplo para datos sin agrupar, y otro para datos agrupados.

5.3.3.1 Ejemplo para datos sin agrupar.

Utilizando los datos de viviendas particulares con techos de materiales ligero, naturales y precarios, de la sección 6.2.3, se calcularán los percentiles 10, 20, 30, 40, 50, 60, 70, 80 y 90.

Para calcular el percentil 10.

Paso 1. Se ordenan los datos de menor a mayor número de viviendas con esa característica (véase Cuadro I.5-10).

CUADRO I.5-10

Viviendas con materiales precarios por delegación en el Distrito Federal*

	Delegación	Viv c/mat prec*
1	Benito Juárez	2,384
2	Cuauhtémoc	4,194
3	Milpa Alta	6,254
4	Cuajimalpa de Morelos	7,241
5	Miguel Hidalgo	8,127
6	Venustiano Carranza	8,728
7	Tláhuac	11,905
8	Coyoacán	12,159
9	La Magdalena Contreras	12,571
10	Azcapotzalco	13,172
11	Iztacalco	13,358
12	Xochimilco	20,008
13	Álvaro Obregón	21,203
14	Tlalpan	24,838
15	Gustavo A. Madero	34,570
16	Iztapalapa	64,833

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000.

**Viviendas particulares con techos de materiales ligeros, naturales y precarios.

Paso 2. Se calcula la frecuencia acumulada para cada dato (véase Cuadro I.5-11). Para ello, la frecuencia es $(1/n)*100\%$ donde, en este caso n es el número de delegaciones, y en general, n es el número de observaciones. A su vez, la frecuencia acumulada es $(i/n)*100\%$ donde, i es el número de observación.

CUADRO I.5-11

Viviendas con materiales precarios por delegación en el Distrito Federal*

	Delegación	Viv c/mat prec*	Frecuencia acumulada
1	Benito Juárez	2,384	6.25%
2	Cuauhtémoc	4,194	12.50%
3	Milpa Alta	6,254	18.75%
4	Cuajimalpa de Morelos	7,241	25.00%
5	Miguel Hidalgo	8,127	31.25%
6	Venustiano Carranza	8,728	37.50%
7	Tláhuac	11,905	43.75%
8	Coyoacán	12,159	50.00%
9	La Magdalena Contreras	12,571	56.25%
10	Azcapotzalco	13,172	62.50%
11	Iztacalco	13,358	68.75%
12	Xochimilco	20,008	75.00%

13	Álvaro Obregón	21,203	81.25%
14	Tlalpan	24,838	87.50%
15	Gustavo A. Madero	34,570	93.75%
16	Iztapalapa	64,833	100.00%

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000.

**Viviendas particulares con techos de materiales ligeros, naturales y precarios.

Paso 3. El percentil 10, C_{10} se debería de encontrar en el valor que acumulara el 10%. Como ningún dato tiene una frecuencia acumulada de 10%, entonces se tomará al punto medio del dato que acumule menos de 10% pero que el siguiente dato acumule más de 10% y el dato que acumule más de 10% pero que el dato anterior acumule menos de 10%. En este caso, será el punto medio del dato 1 y el dato 2, entonces:

$$C_{10} = \frac{1}{2}(Dato1 + Dato2) = \frac{1}{2}(2384 + 4194) = 3289$$

Para calcular el percentil 20

Paso 1. Véase Tabla 10.

Paso 2. Véase Tabla 11.

Paso 3. En este caso, se tomará al punto medio del dato 3 y el dato 4, entonces:

$$C_{20} = \frac{1}{2}(6254 + 7241) = 6747.5$$

Para calcular el percentil 30

Paso 1. Véase Tabla 10.

Paso 2. Véase Tabla 11.

Paso 3. En este caso, se tomará al punto medio del dato 4 y el dato 5, entonces:

$$C_{30} = \frac{1}{2}(7241 + 8127) = 7684$$

Para calcular el percentil 40

Paso 1. Véase Tabla 10.

Paso 2. Véase Tabla 11.

Paso 3. En este caso, se tomará al punto medio del dato 6 y el dato 7, entonces:

$$C_{40} = \frac{1}{2}(8728 + 11905) = 10316.5$$

Para calcular el percentil 50

Paso 1. Véase Tabla 10.

Paso 2. Véase Tabla 11.

Paso 3. En este caso, el percentil 50, C_{50} , corresponde al dato 8, ya que ese dato tiene una frecuencia acumulada de 50%, por lo que $C_{50} = 12159$.

Para calcular el percentil 60

Paso 1. Véase Tabla 10.

Paso 2. Véase Tabla 11.

Paso 3. En este caso, se tomará al punto medio del dato 9 y el dato 10, entonces:

$$C_{60} = \frac{1}{2}(12571 + 13172) = 12871.5$$

Para calcular el percentil 70

Paso 1. Véase Tabla 10.

Paso 2. Véase Tabla 11.

Paso 3. En este caso, se tomará al punto medio del dato 11 y el dato 12, entonces:

$$C_{70} = \frac{1}{2}(13358 + 20008) = 16683$$

Para calcular el percentil 80

Paso 1. Véase Tabla 10.

Paso 2. Véase Tabla 11.

Paso 3. En este caso, se tomará al punto medio del dato 12 y el dato 13, entonces:

$$C_{80} = \frac{1}{2}(20008 + 21203) = 20605.5$$

Para calcular el percentil 90

Paso 1. Véase Tabla 10.

Paso 2. Véase Tabla 11.

Paso 3. En este caso, se tomará al punto medio del dato 14 y el dato 15, entonces:

$$C_{90} = \frac{1}{2}(24838 + 34570) = 29704$$

La interpretación de los percentiles resulta muy sencilla. Para este ejemplo el percentil 30 nos dice que el 30% de las delegaciones tiene menos de 7684 viviendas particulares con techos de materiales ligeros, naturales y precarios, mientras que el 70% de las delegaciones tiene más de 7684 viviendas con esa característica.

5.3.3.2 Ejemplo para datos agrupados.

Utilizando los datos de viviendas particulares con techos de materiales ligeros, naturales y precarios, de la sección 5.3.2, se calcularán los percentiles 30, 50, y 90.

Antes de empezar a calcular los percentiles, calcularemos la frecuencia de cada intervalo, n_i , así como la frecuencia acumulada hasta el intervalo

anterior, n_a (véase Cuadro I.5-12). La frecuencia del intervalo es en este caso, el número de delegaciones que se encuentran en el intervalo. En términos generales, la frecuencia del intervalo es el número de observaciones en ese intervalo. La frecuencia acumulada del intervalo es en este caso, la suma del número de delegaciones de los intervalos anteriores. En general, la frecuencia acumulada del intervalo anterior es la suma del número de observaciones en los intervalos anteriores (siempre y cuando los intervalos estén ordenados de menor a mayor).

CUADRO I.5-12
Delegaciones por cantidad de viviendas con techos ligeros*

Intervalo	n_i^*	n_a
Menos de 10,000 viviendas**	6	0
De 10,000 a 19,999 viviendas**	5	6
De 20,000 a 29,999 viviendas**	3	11
De 30,000 a 70,000 viviendas**	2	14
Total	16	16

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000.

**Se refiere a viviendas particulares con techos de materiales ligeros, naturales y precarios.

Para calcular el percentil 30

Paso 1. Se debe determinar el número de observaciones que representan al 30% de la distribución de frecuencias. En este caso será:

$$(30)\left(\frac{16}{100}\right) = 4.8$$

Paso 2. Comenzando por el intervalo más pequeño se deben calcular las frecuencias acumuladas (véase Cuadro I.5-12). El percentil 30, C_{30} , se encontrará en el intervalo cuya frecuencia acumulada supere el 30% de las observaciones (véase Cuadro I.5-13). En este caso, el primer intervalo es el que cumple esa propiedad, por lo que C_{30} se encontrará en el primer intervalo.

CUADRO I.5-13
Delegaciones por cantidad de viviendas con techos ligeros*

Intervalo	n_i^*	n_a	Frecuencia acumulada	% de frecuencia acumulada
De 0 a 10,000 viviendas**	6	0	6	37.50%
De 10,000 a 19,999 vivienda**s	5	6	11	68.75%
De 20,000 a 29,999 viviendas**	3	11	14	87.50%
De 30,000 a 70,000 viviendas**	2	14	16	100.00%
Total	16	16		

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000.

**Se refiere a viviendas particulares con techos de materiales ligeros, naturales y precarios.

Paso 3. Se aplica la fórmula:

$$C_{30} = 0 + \frac{10000}{6} \left((30) \left(\frac{16}{100} \right) - 0 \right) = 1666.67(4.8) = 8000.$$

Para calcular el percentil 50.

Paso 1. Se debe determinar el número de observaciones que representan al 50% de la distribución de frecuencias. En este caso será:

$$(50) \left(\frac{16}{100} \right) = 8.$$

Paso 2. En este caso el intervalo 2 es el que cumple con la propiedad de que su frecuencia acumulada supera el 50% (véase Tabla 13), por lo que C_{50} se encontrará en el segundo intervalo.

Paso 3. Se aplica la fórmula:

$$C_{50} = 10000 + \frac{9999}{5} \left((50) \left(\frac{16}{100} \right) - 6 \right) = 10000 + 1999.8(8 - 2) = 10000 + 1999.8(2) = 13999.6.$$

Para calcular el percentil 90.

Paso 1. El número de observaciones que representan al 90% de la distribución de frecuencias es:

$$(90) \left(\frac{16}{100} \right) = 14.4$$

Paso 2. En este caso el intervalo 4 es el que cumple con la propiedad de que su frecuencia acumulada supera el 90% (véase Tabla 13), por lo que C_{90} se encontrará en el cuarto intervalo.

Paso 3. Se aplica la fórmula:

$$\begin{aligned} C_{90} &= 30000 + \left(\frac{40000}{2} \right) \left((90) \left(\frac{16}{100} \right) - 14 \right) = 30000 + 20000(14.4 - 14) = 30000 + 20000(0.4) \\ &= 30000 + 8000 = 38000. \end{aligned}$$

La interpretación de los percentiles es idéntica para datos agrupados. Pongamos atención en el percentil 50. De acuerdo a los cálculos para datos agrupados, el 50% de las delegaciones tiene menos de 14,000 viviendas particulares con techos o materiales ligeros, naturales y precarios. Ahora, observando el mismo percentil, pero calculado para datos no agrupados, vemos que su valor es de 12,159. Debido a que el percentil 50 tuvo un cálculo exacto para datos no agrupados, podemos ver que la diferencia existente se

presenta para datos agrupados. Por ello, a menos que sea imposible trabajar con los datos sin agrupar, siempre será preferible su uso.

5.3.4 Los deciles.

Los deciles son 9 valores de una variable que dividen a los datos en 10 partes iguales. Ahora, recordemos que el percentil 10 es el que nos dice que debajo de él se encuentran el 10% de nuestros valores, y encima de él se encuentran el 90% de nuestros valores; también recordemos que el percentil 20 divide a nuestros valores de tal forma que el 20% de los valores se encuentran por debajo de él, y el 80% de los valores se encuentran encima de él. Si seguimos con ese proceso, es fácil notar que los percentiles 10, 20, 30, 40, 50, 60, 70, 80, y 90, son 9 valores que dividen a una variable en 10 partes iguales, que corresponde con la definición de los deciles. Entonces, el decil 1 es el percentil 10; el decil 2 es el percentil 20; el decil 3 es el percentil 30; el decil 4 es el percentil 40, el decil 5 es el percentil 50; el decil 6 es el percentil 60; el decil 7 es el percentil 70; el decil 8 es el percentil 80; y el decil 9 es el percentil 90.

Si bien se pueden calcular los deciles a partir de los percentiles, los pasos para calcular a los deciles se pueden simplificar un poco. Para datos agrupados, los pasos a seguir, una vez ordenados los intervalos de menor a mayor, para calcular el decil k , D_k son:

1. Determinar el número de observaciones que representan el $(k*10)\%$ de la distribución de frecuencias.
2. Calcular las frecuencias acumuladas, comenzando por el intervalo más pequeño. El decil k , D_k , se encontrará en aquel intervalo cuya frecuencia acumulada supere el $(k*10)\%$ de las observaciones. En el caso de que la frecuencia acumulada coincidiera exactamente con el $(k*10)\%$ de las observaciones, C_k sería el límite superior del intervalo.
3. Una vez encontrado el intervalo crítico, aplicar la siguiente fórmula:

$$D_k = L_i + \frac{L}{n_i} \left(k * \frac{n}{10} - n_a \right), \text{ donde,}$$

L_i es el límite exacto inferior del intervalo crítico.

l es la amplitud real del intervalo.

n_i es la frecuencia del intervalo crítico.

n_a es la frecuencia acumulada hasta el intervalo anterior al intervalo crítico.

n es el número total de observaciones.

El cálculo de los deciles para datos sin agrupar es el siguiente:

1. Ordenar los datos de menor a mayor.
2. Calcular la frecuencia acumulada de cada dato.
3. El decil k , D_k se encontrará en el dato que tenga una frecuencia acumulada del $(k*10)\%$. En caso de que ningún dato cumpla con esa característica, se tomara al valor medio del dato que tenga una frecuencia acumulada menor a $(k*10)\%$, pero que el siguiente dato tenga una frecuencia acumulada mayor a $(k*10)\%$ y el dato que tenga una frecuencia acumulada mayor a $(k*10)\%$ pero que el dato anterior tenga una frecuencia acumulada menor a $(k*10)\%$.

5.3.4.1 Ejemplo para datos sin agrupar.

En la Encuesta Nacional de Prácticas y Consumo Cultural, levantada por la Unidad de Estudios Sobre la Opinión (UDESOS) del Instituto de Investigaciones Sociales de la UNAM (IISUNAM) conjuntamente con el Consejo Nacional para la Cultura y las Artes (CONACULTA) durante abril de 2004, se realizó la siguiente pregunta: ¿Cuántos libros tiene en su casa? Los resultados variaron desde las personas que tienen un libro, hasta las personas que en su casa tienen 995 libros (véase Cuadro I.5-14).

CUADRO I.5-14
¿Cuántos libros tiene en su casa?*

¿Cuántos libros tiene en su casa?	Frecuencia	¿Cuántos libros tiene en su casa?	Frecuencia	¿Cuántos libros tiene en su casa?	Frecuencia
1	78	51	1	148	1
2	90	52	3	150	57
3	96	53	2	151	2
4	84	54	1	152	2
5	160	55	2	160	2
6	75	57	2	161	1
7	54	58	3	170	2
8	71	59	2	180	7
9	18	60	82	185	1
10	339	62	2	194	1
11	14	63	2	195	1
12	57	64	1	197	1
13	18	65	5	200	100
14	15	70	47	211	1
15	181	73	1	220	1
16	20	75	3	221	1
17	10	76	1	222	1
18	22	77	1	240	1
19	4	80	63	250	10
20	304	82	2	251	1
21	5	85	1	270	1
22	11	87	1	300	62
23	6	88	1	301	1
24	10	90	14	302	2
25	65	93	1	311	1
26	4	95	4	350	5
27	7	96	2	400	12
28	12	97	1	401	2
29	1	98	2	402	2
30	232	99	2	450	1
32	9	100	179	480	1
33	3	101	4	490	1
34	2	102	5	500	17
35	28	108	1	502	2
36	5	110	2	522	2
37	2	111	1	600	4
38	7	112	3	700	2
39	1	113	2	711	2
40	152	115	3	752	1
42	5	118	1	796	1
43	1	120	9	800	4
44	1	121	5	804	1
45	20	122	3	821	1
46	1	125	1	900	2
48	4	126	1	901	1
49	1	130	5	990	2

50	291	140	2	995	1
----	-----	-----	---	-----	---

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Prácticas y Consumo Cultural, 2004. UDES, IISUNAM, CONACULTA.

En primer lugar, hay que resaltar que estos son datos sin agrupar. En efecto, por ejemplo, el tabulado presenta de forma resumida que existen 232 personas encuestadas que dijeron tener 30 libros en su casa (véase Cuadro I.5-14), pero no representa a un intervalo, sino una forma de no repetir el dato de 30 libros 232 veces.

Con esos datos, vamos a calcular los deciles 3, 6, y 9.

Paso 1. Los datos presentados en el Cuadro I.5-14 ya están ordenados de menor a mayor.

Paso 2. Existen 3,406 valores. Por ello, si ordenamos los datos de menor a mayor, y los valores son $x_1, \dots, x_i, \dots, x_{3406}$, la frecuencia acumulada del dato i -ésimo sigue la fórmula $i/3406$. Sería larguísima una tabla que presentara todos esos valores, por ello, se construirán las frecuencias acumuladas a partir del Cuadro I.5-14. La idea es la siguiente: Para la frecuencia i -ésima, y_i , la frecuencia acumulada será: $y_1 + y_2 + \dots + y_{i-2} + y_{i-1} + y_i$ (véase Cuadro I.5-15). Y esos valores representan a la frecuencia del último individuo colocado en la escala, que tiene esa cantidad de libros.

CUADRO I.5-15
¿Cuántos libros tiene en su casa?*

¿Cuántos libros tiene en su casa?	Frecuencia	Frecuencia Acumulada	¿Cuántos libros tiene en su casa?	Frecuencia	Frecuencia Acumulada	¿Cuántos libros tiene en su casa?	Frecuencia	Frecuencia Acumulada
1	78	2.29%	51	1	76.25%	148	1	90.40%
2	90	4.93%	52	3	76.34%	150	57	92.07%
3	96	7.75%	53	2	76.39%	151	2	92.13%
4	84	10.22%	54	1	76.42%	152	2	92.19%
5	160	14.91%	55	2	76.48%	160	2	92.25%
6	75	17.12%	57	2	76.54%	161	1	92.28%
7	54	18.70%	58	3	76.63%	170	2	92.34%
8	71	20.79%	59	2	76.69%	180	7	92.54%
9	18	21.32%	60	82	79.10%	185	1	92.57%
10	339	31.27%	62	2	79.15%	194	1	92.60%
11	14	31.68%	63	2	79.21%	195	1	92.63%
12	57	33.35%	64	1	79.24%	197	1	92.66%

13	18	33.88%	65	5	79.39%	200	100	95.60%
14	15	34.32%	70	47	80.77%	211	1	95.63%
15	181	39.64%	73	1	80.80%	220	1	95.65%
16	20	40.22%	75	3	80.89%	221	1	95.68%
17	10	40.52%	76	1	80.92%	222	1	95.71%
18	22	41.16%	77	1	80.95%	240	1	95.74%
19	4	41.28%	80	63	82.80%	250	10	96.04%
20	304	50.21%	82	2	82.85%	251	1	96.07%
21	5	50.35%	85	1	82.88%	270	1	96.10%
22	11	50.68%	87	1	82.91%	300	62	97.92%
23	6	50.85%	88	1	82.94%	301	1	97.94%
24	10	51.15%	90	14	83.35%	302	2	98.00%
25	65	53.05%	93	1	83.38%	311	1	98.03%
26	4	53.17%	95	4	83.50%	350	5	98.18%
27	7	53.38%	96	2	83.56%	400	12	98.53%
28	12	53.73%	97	1	83.59%	401	2	98.59%
29	1	53.76%	98	2	83.65%	402	2	98.65%
30	232	60.57%	99	2	83.71%	450	1	98.68%
32	9	60.83%	100	179	88.96%	480	1	98.71%
33	3	60.92%	101	4	89.08%	490	1	98.74%
34	2	60.98%	102	5	89.22%	500	17	99.24%
35	28	61.80%	108	1	89.25%	502	2	99.30%
36	5	61.95%	110	2	89.31%	522	2	99.35%
37	2	62.01%	111	1	89.34%	600	4	99.47%
38	7	62.21%	112	3	89.43%	700	2	99.53%
39	1	62.24%	113	2	89.49%	711	2	99.59%
40	152	66.71%	115	3	89.58%	752	1	99.62%
42	5	66.85%	118	1	89.61%	796	1	99.65%
43	1	66.88%	120	9	89.87%	800	4	99.77%
44	1	66.91%	121	5	90.02%	804	1	99.79%
45	20	67.50%	122	3	90.11%	821	1	99.82%
46	1	67.53%	125	1	90.14%	900	2	99.88%
48	4	67.65%	126	1	90.16%	901	1	99.91%
49	1	67.67%	130	5	90.31%	990	2	99.97%
50	291	76.22%	140	2	90.37%	995	1	100.00%

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Prácticas y Consumo Cultural, 2004. UDESOL, IISUNAM, CONACULTA.

Paso 3a. Para calcular el decil 3, nos fijamos que para las personas que tienen 10 libros en su casa es en donde se acumula la primera frecuencia mayor al 30% (véase Cuadro I.5-15). La frecuencia para ese caso es de 339, eso quiere decir que alguna de esas 339 personas es la primera en tener una frecuencia mayor al 30%, y la anterior tiene una frecuencia acumulada inferior al 30%. Como para ambas personas el número de libros en su casa es de 10, el valor medio es de 10, y entonces $D_3 = 10$.

Paso 3b. Para calcular el decil 6, nos fijamos que para las personas que tienen 30 libros en su casa es en donde se acumula la primera frecuencia mayor al 60% (véase Cuadro I.5-15). La frecuencia para ese caso es de 232, eso quiere decir que alguna de esas 232 personas es la primera en tener una frecuencia mayor al 60%, y la anterior tiene una frecuencia acumulada inferior al 60%. Como para ambas personas el número de libros en su casa es de 30, el valor medio es de 30, y entonces $D_6 = 30$.

Paso 3c. Para calcular el decil 9, nos fijamos que para las personas que tienen 121 libros en su casa es en donde se acumula la primera frecuencia mayor al 90% (véase Cuadro I.5-15). La frecuencia para ese caso es de 5, eso quiere decir que alguna de esas 5 personas es la primera en tener una frecuencia mayor al 90%, y la anterior tiene una frecuencia acumulada inferior al 90%. Como para ambas personas el número de libros en su casa es de 121, el valor medio es de 121, y entonces $D_9 = 121$.

La interpretación de los deciles es idéntica a la de los percentiles. En este caso, tenemos que el 30% de las personas encuestadas tiene 10 libros o menos en su casa. Además, el 60% de las personas encuestadas tiene 30 libros o menos en su casa, y por último, el 90% de las personas encuestadas tiene 121 libros o menos en su casa.

Existe una manera alternativa para calcular los percentiles en este caso.

Fijémonos en que la frecuencia acumulada de 100% se alcanza en el dato 3406. En otras palabras, hubo 3406 personas que contestaron esta pregunta. Cada persona acumula $(100/3406)\% = 0.0294\%$. Esto quiere decir que si ordenamos a las 3406 personas, la persona i -ésima tendrá una frecuencia acumulada de $(i*100/3406)\%$. Siguiendo ese razonamiento, tenemos:

$$f_i = i \left(\frac{100}{3406} \right) \% \Rightarrow i = \frac{f_i}{\left(\frac{100}{3406} \right) \%}$$

Entonces, si queremos calcular D_3 , el decil 3, se tiene:

$$i = 30 / (100 / 3406) = 30 / (0.0293599) = 1021.8$$

Y esto significa que el decil 3 se alcanzaría en la persona 1,021.8. Como no hay fracción de persona, esto significa que el decil 3 es igual al valor medio de la persona 1,021 y la persona 1,022. Si en el Cuadro I.5-15, en lugar de expresar a las frecuencias acumuladas en porcentajes, las expresamos como el número de personas que han dicho tener esa cantidad de libros, o menos, tendremos una frecuencia absoluta acumulada. A la que se expresa como porcentajes se le llama relativa, porque es relativa al número total de personas, y a la que expresa número de personas se le llama absoluta porque en todo momento expresa que nos referimos a personas (véase Cuadro I.5-16).

CUADRO I.5-16
¿Cuántos libros tiene en su casa?*

¿Cuántos libros tiene en su casa?	Frecuencia	Frecuencia Absoluta Acumulada	¿Cuántos libros tiene en su casa?	Frecuencia	Frecuencia Acumulada	¿Cuántos libros tiene en su casa?	Frecuencia	Frecuencia Acumulada
1	78	78	51	1	2,597	148	1	3,079
2	90	168	52	3	2,600	150	57	3,136
3	96	264	53	2	2,602	151	2	3,138
4	84	348	54	1	2,603	152	2	3,140
5	160	508	55	2	2,605	160	2	3,142
6	75	583	57	2	2,607	161	1	3,143
7	54	637	58	3	2,610	170	2	3,145
8	71	708	59	2	2,612	180	7	3,152
9	18	726	60	82	2,694	185	1	3,153
10	339	1,065	62	2	2,696	194	1	3,154
11	14	1,079	63	2	2,698	195	1	3,155
12	57	1,136	64	1	2,699	197	1	3,156
13	18	1,154	65	5	2,704	200	100	3,256
14	15	1,169	70	47	2,751	211	1	3,257
15	181	1,350	73	1	2,752	220	1	3,258
16	20	1,370	75	3	2,755	221	1	3,259
17	10	1,380	76	1	2,756	222	1	3,260
18	22	1,402	77	1	2,757	240	1	3,261
19	4	1,406	80	63	2,820	250	10	3,271
20	304	1,710	82	2	2,822	251	1	3,272
21	5	1,715	85	1	2,823	270	1	3,273
22	11	1,726	87	1	2,824	300	62	3,335
23	6	1,732	88	1	2,825	301	1	3,336
24	10	1,742	90	14	2,839	302	2	3,338
25	65	1,807	93	1	2,840	311	1	3,339
26	4	1,811	95	4	2,844	350	5	3,344
27	7	1,818	96	2	2,846	400	12	3,356

28	12	1,830	97	1	2,847	401	2	3,358
29	1	1,831	98	2	2,849	402	2	3,360
30	232	2,063	99	2	2,851	450	1	3,361
32	9	2,072	100	179	3,030	480	1	3,362
33	3	2,075	101	4	3,034	490	1	3,363
34	2	2,077	102	5	3,039	500	17	3,380
35	28	2,105	108	1	3,040	502	2	3,382
36	5	2,110	110	2	3,042	522	2	3,384
37	2	2,112	111	1	3,043	600	4	3,388
38	7	2,119	112	3	3,046	700	2	3,390
39	1	2,120	113	2	3,048	711	2	3,392
40	152	2,272	115	3	3,051	752	1	3,393
42	5	2,277	118	1	3,052	796	1	3,394
43	1	2,278	120	9	3,061	800	4	3,398
44	1	2,279	121	5	3,066	804	1	3,399
45	20	2,299	122	3	3,069	821	1	3,400
46	1	2,300	125	1	3,070	900	2	3,402
48	4	2,304	126	1	3,071	901	1	3,403
49	1	2,305	130	5	3,076	990	2	3,405
50	291	2,596	140	2	3,078	995	1	3,406

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Prácticas y Consumo Cultural, 2004. UDES, IISUNAM, CONACULTA.

En el Cuadro I.5-16 podemos apreciar que las personas 1,021 y 1,022 tienen 10 libros en su casa, ya que, al ser 726 las personas que tienen 9 libros o menos en su casa, y ser 339 las personas que tienen 10 libros en su casa, si las ordenamos de menor a mayor número de libros en su casa, las personas 1,021 y 1,022 no tienen 9 libros o menos (porque según el ordenamiento, sólo las primeras 726 personas cumplen con esa propiedad), por lo que tienen 10 libros, y por lo tanto, el valor medio es 10. Por lo anterior, se tiene que $D_3 = 10$, lo que significa que el 30% de la población encuestada tiene 10 libros o menos en su casa.

Siguiendo con esa lógica,

$$i = 60 / 0.0293599 = 2043.6, \text{ y}$$

$$i = 90 / 0.0293599 = 3065.4.$$

Por lo que el decil 6 se encontrará en el valor medio de las personas 2,043 y 2,044. Consecuentemente, el decil 9 se encontrará en el valor medio de las personas 3,065 y 3,066. Fácilmente podemos ver que tanto las personas 2,043 y 2,044 tienen 30 libros, y que las personas 3,065 y 3,066 tienen 121 libros. Por lo que, $D_6 = 30$ y $D_9 = 121$. Y llegamos a las mismas conclusiones

que con el otro método: el 60% de las personas encuestadas tienen 30 libros o menos en su casa, y el 90% de las personas encuestadas tienen 121 libros o menos en su casa.

5.3.4.2 Ejemplo para datos agrupados.

En este ejemplo, consideraremos a la misma encuesta que en la sección anterior. Se calcularán los deciles 4, 8 y 9. Para ello, se utilizarán a los datos agrupados en 11 grupos (véase Cuadro I.5-17).

CUADRO I.5-16
¿Cuántos libros tiene en su casa?*

	¿Cuántos libros tiene en casa?	n_i	n_a	Frecuencia acumulada	% de frecuencia acumulada
1	De 1 a 10	1,045	0	1,045	30.68%
2	De 11 a 20	653	1,045	1,698	49.85%
3	De 21 a 30	363	1,698	2,061	60.51%
4	De 31 a 40	211	2,061	2,272	66.71%
5	De 41 a 50	315	2,272	2,587	75.95%
6	De 51 a 60	93	2,587	2,680	78.68%
7	De 61 a 70	57	2,680	2,737	80.36%
8	De 71 a 80	74	2,737	2,811	82.53%
9	De 81 a 90	18	2,811	2,829	83.06%
10	De 91 a 100	210	2,829	3,039	89.22%
11	De 101 a 1,000	367	3,039	3,406	100.00%
	Total	3,406	3,406		

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Prácticas y Consumo Cultural, 2004. UDES0, IISUNAM, CONACULTA.

Para el decil 4.

Paso 1. El número de observaciones que representa al 40% de la distribución de frecuencias es: $40\left(\frac{3406}{100}\right) = 1362.4$

Paso 2. Es claro que el decil 4 se encuentra en el segundo intervalo (ya que es el primer intervalo en acumular más del 40%).

Paso 3.

$$D_4 = 10 + \frac{9}{653} \left(4 * \frac{3406}{10} - 1045 \right) = 10 + 0.0137825(1362.4 - 1045) = 10 + 0.0137825(317.4) = 10 + 4.37 = 14.37$$

Para el decil 8.

Paso 1. El número de observaciones que representa al 80% de la distribución de frecuencias es: $80\left(\frac{3406}{100}\right) = 2724.8$.

Paso 2. Es claro que el decil 8 se encuentra en el intervalo 7 (ya que es el primer intervalo que acumula más del 80%).

Paso 3.

$$D_8 = 61 + \frac{9}{57}(2724.8 - 2680) = 61 + 0.1579(44.8) = 61 + 7.07 = 68.07 .$$

Para el decil 9.

Paso 1. El número de observaciones que representa al 90% de la distribución de frecuencias es: $90\left(\frac{3406}{100}\right) = 3065.4$.

Paso 2. El decil 9 se encuentra en el intervalo 11, ya que es el primero en superar al 90% de la distribución de frecuencias.

Paso 3.

$$D_9 = 101 + \frac{899}{367}(3065.4 - 3039) = 101 + 2.4496(26.4) = 101 + 64.6692 = 165.6692 .$$

Del cálculo de los deciles para datos agrupados se puede concluir que el 40% de la población encuestada tiene 15 libros o menos en su casa (debido a que no hay fracciones de libros, se considera al siguiente número a menos que el cálculo sea exacto), el 80% de la población encuestada tiene 69 libros o menos, y el 90% de la población encuestada tiene 166 libros o menos. Si los comparamos con los datos sin agrupar (los cálculos son sencillos y se deja al lector verificarlos), se puede ver que el cálculo fue exacto para el decil 4; se subestimó para el decil 8; y se sobreestimó para el decil 9.

5.3.5 Los cuartiles.

Los cuartiles son 3 valores de una variable que dividen a los datos en 4 partes iguales. Ahora, recordemos que el percentil 25 es el que nos dice que debajo de él se encuentran el 25% de nuestros valores, y encima de él se encuentran el 75% de nuestros valores; también recordemos que el percentil 50 divide a nuestros valores de tal forma que el 50% de los valores se encuentran por debajo de él, y el 50% de los valores se encuentran encima de él. Si seguimos con ese proceso, es fácil notar que los percentiles 25, 50, y 75, son 3 valores que dividen a una variable en 4 partes iguales, lo que corresponde con la definición de un cuartil. Entonces, el cuartil 1 es el percentil 25; el cuartil 2 es el percentil 50; y el cuartil 3 es el percentil 75.

Si bien se pueden calcular los cuartiles a partir de los percentiles, los pasos para calcular a los cuartiles se pueden simplificar un poco. Para datos agrupados, los pasos a seguir, una vez ordenados los intervalos de menor a mayor, para calcular el Cuartil k , Q_k son:

1. Determinar el número de observaciones que representan el $(k*25)\%$ de la distribución de frecuencias.
2. Calcular las frecuencias acumuladas, comenzando por el intervalo más pequeño. El cuartil k , Q_k , se encontrará en aquel intervalo cuya frecuencia acumulada supere el $(k*25)\%$ de las observaciones. En el caso de que la frecuencia acumulada coincidiera exactamente con el $(k*10)\%$ de las observaciones, C_k sería el límite superior del intervalo.
3. Una vez encontrado el intervalo crítico, aplicar la siguiente fórmula:

$$Q_k = L_i + \frac{l}{n_i} (k * \frac{n}{4} - n_a), \text{ donde,}$$

L_i es el límite exacto inferior del intervalo crítico.

l es la amplitud real del intervalo.

n_i es la frecuencia del intervalo crítico.

n_a es la frecuencia acumulada hasta el intervalo anterior al intervalo crítico.

n es el número total de observaciones.

El cálculo de los cuartiles para datos sin agrupar es el siguiente:

1. Ordenar los datos de menor a mayor.
2. Calcular la frecuencia acumulada de cada dato.
3. El cuartil k , Q_k se encontrará en el dato que tenga una frecuencia acumulada del $(k*25)\%$. En caso de que ningún dato cumpla con esa característica, se tomara al valor medio del dato que tenga una frecuencia acumulada menor a $(k*25)\%$, pero que el siguiente dato tenga una frecuencia acumulada mayor a $(k*25)\%$ y el dato que tenga una frecuencia acumulada mayor a $(k*25)\%$ pero que el dato anterior tenga una frecuencia acumulada menor a $(k*25)\%$.

Como se puede apreciar, el cálculo de los cuartiles resulta idéntico al de los percentiles. Veremos un ejemplo más.

5.3.5.1 Ejemplo.

El INEGI, en su censo, también levantó un cuestionario ampliado. En él, se tiene la información de las 32 entidades federativas del país. Para este ejemplo, se trabajará con todas las personas encuestadas que trabajaron en el estado de Oaxaca. En total fueron encuestadas 218,958 personas, de las cuales 209,146 especificaron su ingreso, y de ellos, 113,894 reciben algún ingreso. Por la cantidad de datos que se tiene resulta imposible presentar los datos sin agrupar, por ello, se trabajará con datos agrupados. En el periodo del levantamiento, el salario mínimo para Oaxaca fue de \$981.00 diarios³. Se desea saber cómo se distribuye el ingreso de acuerdo a la población. Para ello, se agrupará a los datos de acuerdo al salario mínimo (véase Cuadro I.5-18), y se calcularán los cuartiles 1, 2, y 3. También se calcularán el decil 9, y los percentiles 99, y 99.5.

³ FUENTE: Comisión Nacional de los Salarios Mínimos. Salario mínimo para el área C (que comprende a todo el estado de Oaxaca).

CUADRO I.5-18
Distribución del ingreso entre los habitantes de Oaxaca*

	Cifras en pesos	Cifras en salarios mínimos	n_i	n_a	Frecuencia acumulada	% de frecuencia acumulada
1	Más de \$0 a menos de \$245.25	0 - 1/4 SM	9,047	0	9,047	7.94%
2	Más de \$245.25 a menos de \$490.50	1/4 - 1/2 SM	10,444	9,047	19,491	17.11%
3	Más de \$490.50 a menos de \$981	1/2 - 1 SM	26,178	19,491	45,669	40.10%
4	Más de \$981 a menos de \$1,962	1 - 2 SM	36,452	45,669	82,121	72.10%
5	Más de \$1,962 a menos de \$2,943	2 - 3 SM	13,331	82,121	95,452	83.81%
6	Más de \$2,943 a menos de \$3,924	3 - 4 SM	7,790	95,452	103,242	90.65%
7	Más de \$3,924 a menos de \$4,905	4 - 5 SM	5,008	103,242	108,250	95.04%
8	Más de \$4,905 a menos de \$5,886	5 - 6 SM	1,734	108,250	109,984	96.57%
9	Más de \$5,886 a menos de \$6,867	6 - 7 SM	1,303	109,984	111,287	97.71%
10	Más de \$6,867 a menos de \$7,848	7 - 8 SM	410	111,287	111,697	98.07%
11	Más de \$7,848 a menos de \$8,829	8 - 9 SM	697	111,697	112,394	98.68%
12	Más de \$8,829 a menos de \$9,810	9 - 10 SM	141	112,394	112,535	98.81%
13	Más de \$9,810 a menos de \$14,715	10 - 15 SM	746	112,535	113,281	99.46%
14	Más de \$14,715 a menos de \$19,620	15 - 20 SM	238	113,281	113,519	99.67%
15	Más de \$19,620 a menos de \$29,430	20 - 30 SM	188	113,519	113,707	99.84%
16	Más de \$29,430 a menos de \$49,050	30 - 50 SM	99	113,707	113,806	99.92%
17	Más de \$49,050 a menos de \$98,100	50 - 100 SM	20	113,806	113,826	99.94%
18	Más de \$98,100 a menos de \$1,000,000	100SM o más	68	113,826	113,894	100.00%
		Total	113,894	113,894		

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: XII Censo General de Población y Vivienda. Cuestionario Ampliado. INEGI, 2000. Se refiere al ingreso percibido por cada habitante que percibió algún ingreso.

Para el cuartil 1.

Paso 1. El número de observaciones que representa al 25% de la distribución de frecuencias es: $25\left(\frac{113894}{100}\right) = 28473.5$.

Paso 2. Q_1 se encuentra en el intervalo 3.

Paso 3. Aplicando la fórmula:

$$Q_1 = 490.50 + \frac{490.50}{26178} (28473.5 - 19491) = 490.50 + (0.0187371)(8982.5) = 490.50 + 163.306 = 658.806 = \$658.81$$

Para el cuartil 2.

Paso 1. El número de observaciones que representa al 50% de la distribución de frecuencias es: $50\left(\frac{113894}{100}\right) = 56947$.

Paso 2. Q_2 se encuentra en el intervalo 4.

Paso 3. Aplicando la fórmula:

$$Q_2 = 981 + \frac{981}{36452} (56947 - 45669) = 981 + (0.0269121)(11278) = 981 + 303.51466 = 1284.5146 = \$1284.51$$

Para el cuartil 3.

Paso 1. El número de observaciones que representa al 75% de la distribución de frecuencias es: $75\left(\frac{113894}{100}\right) = 85420.5$.

Paso 2. Q_3 se encuentra en el intervalo 5.

Paso 3. Aplicando la fórmula:

$$Q_3 = 1962 + \frac{981}{13331}(85420.5 - 82121) = 1962 + (0.0735878)(3299.5) = 1962 + 242.80294 = 2204.8029 \\ = \$2204.80$$

Para el decil 9.

Paso 1. El número de observaciones que representa al 90% de la distribución de frecuencias es: $90\left(\frac{113894}{100}\right) = 102504.6$.

Paso 2. D_9 se encuentra en el intervalo 6.

Paso 3. Aplicando la fórmula:

$$D_9 = 2943 + \frac{981}{7790}(102504.6 - 95452) = 2943 + (0.1259306)(7052.6) = 2943 + 888.13814 = 3831.1381 \\ = \$3831.14$$

Para el percentil 99.

Paso 1. El número de observaciones que representa al 99% de la distribución de frecuencias es: $99\left(\frac{113894}{100}\right) = 112755.06$.

Paso 2. C_{99} se encuentra en el intervalo 13.

Paso 3. Aplicando la fórmula:

$$C_{99} = 9810 + \frac{4905}{746}(112755.06 - 112535) = 9810 + (6.575067)(220.06) = 9810 + 1446.9092 = 11256.909 \\ = \$11256.91$$

Para el percentil 99.5.

Paso 1. El número de observaciones que representa al 99.5% de la distribución de frecuencias es: $99.5\left(\frac{113894}{100}\right) = 113324.53$.

Paso 2. $C_{99.5}$ se encuentra en el intervalo 14.

Paso 3. Aplicando la fórmula:

$$C_{99.5} = 14715 + \frac{4905}{238}(113324.53 - 113281) = 14715 + (20.60924)(43.53) = 14715 + 897.1202 = 15612.12 \\ = \$15612.12$$

La interpretación de los cuartiles es igual a la de los deciles, y a la de los percentiles. En este ejemplo, podemos apreciar que el 25% de la población trabajadora que recibe ingresos gana menos de un salario mínimo. De igual forma, se puede ver que el 50% de la población que recibe ingresos gana menos de dos salarios mínimos. Bajo el mismo esquema, el 75% de la población trabajadora que recibe ingresos percibe menos de tres salarios mínimos. El 90% de la población trabajadora que recibe ingresos recibe menos de cuatro salarios mínimos. Otra forma de ver esta información es que el 10% de la población trabajadora que recibe ingresos gana más de 4 salarios mínimos. Desde ese punto de vista, y considerando al percentil 99 se dice que el 1% de la población trabajadora que recibe ingresos percibe más de 10 salarios mínimos, y que el 0.50% de la población trabajadora que recibe ingresos gana más de 15 salarios mínimos.

5.3.6 Conclusión

Las medidas de posición nos permiten describir la manera en que se distribuye la población objetivo en base a una variable dada. Es posible con estas medidas poder decir en dónde se agrupa un porcentaje de los datos, o bien, en donde no se agrupa. Las medidas de posición son importantes, por muchas causas. Una de ellas es que nos permiten observar en dónde se agrupa la mayor parte de los datos. En el ejemplo 5.3.5.1 se apreció que sólo el 1% de la población gana más de 10 salarios mínimos. Si bien en este caso parece normal el resultado, este tipo de valores atípicos también podrían indicar la existencia de un error, por ejemplo, se puede tener un error de captura en el que en lugar de capturar 1000, se capturó 10000, y esa clase de errores que nos pueden afectar los resultados.

5.4 Medidas de Tendencia Central.

Las medidas de tendencia central son estadísticas que nos proporcionan información sobre la posición media de los datos respecto a la distribución de valores de una variable. Las principales estadísticas de tendencia central son la media, la moda, y la mediana, y a continuación se presentan.

5.4.1 La media.

La media de una variable es la suma de todos los valores de dicha variable, dividida por el número total de observaciones. Es decir:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_N}{N}, \text{ donde}$$

X_i representa a la observación i -ésima.

N representa al número total de observaciones, y,

\bar{X} representa a la media, y se lee "equis barra".

La fórmula anterior se aplica a datos sin agrupar, sin embargo, si los datos están agrupados, la fórmula para el cálculo de la media es:

$$\bar{X} = \frac{X_1 n_1 + X_2 n_2 + \dots + X_i n_i + \dots + X_N n_N}{n_1 + n_2 + \dots + n_i + \dots + n_N}, \text{ donde,}$$

X_i representa al valor medio del intervalo i -ésimo.

n_i representa a la frecuencia del intervalo i -ésimo.

A continuación se presentan algunos ejemplos.

5.4.1.1 Ejemplo 1.

Se calculará la media para el ejemplo de la sección 6.3.3.1.

En el Cuadro I.5-10 aparece el número de viviendas particulares con techos de materiales ligeros, naturales y precarios. En este caso, la media será la suma del número de viviendas con esa característica por delegación, entre 16 (el número total de delegaciones). Entonces:

$$\begin{aligned} \bar{X} &= \frac{\left(2384 + 4194 + 6254 + 7241 + 8127 + 8728 + 11905 + 12159 + \right. \\ &\quad \left. 12571 + 13172 + 13358 + 20008 + 21203 + 24838 + 34570 + 64833 \right)}{16} \\ &= \frac{265545}{16} = 16596.56 \end{aligned}$$

Entonces, se dice, que en promedio, cada delegación tiene 16,597 viviendas particulares con techos de materiales ligeros, naturales y precarios.

5.4.1.2 Ejemplo 2.

Se calculará la media para el ejemplo 6.3.3.2.

En este caso, se trata de datos agrupados. Para calcular la media se necesitan calcular los valores medios de cada intervalo, así como el producto del valor medio de cada intervalo y su frecuencia. En el Cuadro I.5-19 se muestran estos valores.

CUADRO I.5-19
Delegaciones por cantidad de viviendas con techos ligeros*

Intervalo	n_i^*	X_i	$X_i n_i$
Menos de 10,000 viviendas**	6	5000	30000
De 10,000 a 19,999 vivienda**s	5	14999.5	74997.5
De 20,000 a 29,999 viviendas**	3	24999.5	74998.5
De 30,000 a 70,000 viviendas**	2	50000	100000
Total	16		

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000.

**Se refiere a viviendas particulares con techos de materiales ligeros, naturales y precarios.

Entonces, aplicando la fórmula para el cálculo de la media para datos agrupados:

$$\begin{aligned} \bar{X} &= \frac{(5000 * 6) + (14999.5 * 5) + (24999.5 * 3) + (50000 * 2)}{6 + 5 + 3 + 2} \\ &= \frac{30000 + 74997.5 + 74998.5 + 100000}{16} = \frac{279996}{16} = 17499.75 \end{aligned}$$

Y, se dice que, en promedio, cada delegación tiene 17,500 viviendas particulares con techos de materiales ligeros, naturales y precarios. Nótese, como en este caso, el cálculo de la media para datos agrupados sobreestimó a la media en poco más de 900 viviendas.

5.4.1.3 Ejemplo 3.

Se calculará la media para el ejemplo 5.3.3.1.

En este caso, el Cuadro I.5-14 muestra la cantidad de personas que tiene un número de libros dado en su casa. El cálculo de la media es en este caso la suma del producto del número de libros por el número de personas que dijo tener ese número de libros en su casa, dividido entre el número total de personas entrevistadas. A continuación se presentan algunos casos:

$$\bar{X} = \frac{\left[\begin{array}{l} (1\text{libro})(78\text{ personas}) + (2\text{libros})(90\text{ personas}) + (3\text{libros})(96\text{ personas}) + \\ \vdots \\ +(901\text{libros})(1\text{ persona}) + (990\text{libros})(2\text{ personas}) + (995\text{libros})(1\text{ persona}) \end{array} \right]}{3406\text{ personas}}$$
$$= \frac{186820}{3406} = 54.85,$$

Y se dice que, en promedio, cada persona tiene 55 libros en su casa.

5.4.1.4 Ejemplo 4.

Se calculará la media para el ejemplo 5.3.4.2.

En este ejemplo se trabaja con datos agrupados. Para calcular la media se necesita calcular los valores medios de cada intervalo, así como el producto del valor medio de cada intervalo y su frecuencia. En el Cuadro I.5-20 se muestran estos valores.

CUADRO I.5-20
¿Cuántos libros tiene en su casa?*

	¿Cuántos libros tiene en casa?	n_i	X_i	$X_i n_i$
1	De 1 a 10	1,045	5.5	5,748
2	De 11 a 20	653	15.5	10,122
3	De 21 a 30	363	25.5	9,257
4	De 31 a 40	211	35.5	7,491
5	De 41 a 50	315	45.5	14,333
6	De 51 a 60	93	55.5	5,162
7	De 61 a 70	57	65.5	3,734
8	De 71 a 80	74	75.5	5,587
9	De 81 a 90	18	85.5	1,539
10	De 91 a 100	210	95.5	20,055
11	De 101 a 1,000	367	550.5	202,034
	Total	3,406	3,406	

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Prácticas y Consumo Cultural, 2004. UDES, IISUNAM, CONACULTA.

Con estos datos se puede aplicar la fórmula para el cálculo de la media para datos agrupados.

$$\bar{X} = \frac{\left(\begin{array}{l} 5748 + 10122 + 9257 + 7491 + 14333 + 5162 \\ + 3734 + 5587 + 1539 + 20055 + 202034 \end{array} \right)}{3406} = \frac{285058}{3406} = 83.69.$$

Y se dice que, en promedio, cada persona tiene 84 libros en su casa. Hay que destacar que en este caso, la media calculada con datos agrupados sobreestimó en casi 30 el número de libros promedio en una casa. En este caso, la sobreestimación fue de más del 50%. Por ello, se debe tener cuidado al momento de trabajar con datos agrupados.

5.4.1.5. Ejemplo 5

Se calculará la media para el ejemplo 5.3.5.1.

En este caso, se trata de datos agrupados. Para calcular la media se necesitan calcular los valores medios de cada intervalo, así como el producto del valor medio de cada intervalo y su frecuencia. En el Cuadro I.5-21 se muestran estos valores.

**CUADRO I.5-21
Nivel de ingreso de las personas en Oaxaca***

Cifras en pesos	Cifras en salarios mínimos	n_i	X_i	$X_i n_i$
Más de \$0 a menos de \$245.25	0 - 1/4 SM	9,047	122.625	1,109,388
Más de \$245.25 a menos de \$490.50	1/4 - 1/2 SM	10,444	367.875	3,842,087
Más de \$490.50 a menos de \$981	1/2 - 1 SM	26,178	735.750	19,260,464
Más de \$981 a menos de \$1,962	1 - 2 SM	36,452	1,471.500	53,639,118
Más de \$1,962 a menos de \$2,943	2 - 3 SM	13,331	2,452.500	32,694,278
Más de \$2,943 a menos de \$3,924	3 - 4 SM	7,790	3,433.500	26,746,965
Más de \$3,924 a menos de \$4,905	4 - 5 SM	5,008	4,414.500	22,107,816
Más de \$4,905 a menos de \$5,886	5 - 6 SM	1,734	5,395.500	9,355,797
Más de \$5,886 a menos de \$6,867	6 - 7 SM	1,303	6,376.500	8,308,580
Más de \$6,867 a menos de \$7,848	7 - 8 SM	410	7,357.500	3,016,575
Más de \$7,848 a menos de \$8,829	8 - 9 SM	697	8,338.500	5,811,935
Más de \$8,829 a menos de \$9,810	9 - 10 SM	141	9,319.500	1,314,050
Más de \$9,810 a menos de \$14,715	10 - 15 SM	746	12,267.500	9,151,555
Más de \$14,715 a menos de \$19,620	15 - 20 SM	238	17,167.500	4,085,865
Más de \$19,620 a menos de \$29,430	20 - 30 SM	188	24,525.000	4,610,700
Más de \$29,430 a menos de \$49,050	30 - 50 SM	99	39,240.000	3,884,760
Más de \$49,050 a menos de \$98,100	50 - 100 SM	20	73,575.000	1,471,500
Más de \$98,100 a menos de \$1,000,000	100SM o más	68	549,050.000	37,335,400
	Total	113,894	113,894	247,746,830

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: XII Censo General de Población y Vivienda. Cuestionario Ampliado. INEGI, 2000.

Entonces, el promedio calculado con la fórmula para datos agrupados es:

$$\bar{X} = \frac{247,746,830}{113,894} = 2,175.2404$$

Y se dice, que en promedio, las personas que trabajan, obtienen un ingreso de \$2,175.24.

Para concluir, hagámonos las siguientes preguntas:

- ¿Realmente se puede decir que una persona tiene 55 libras en promedio en su casa, cuando se vio que más del 75% de las personas encuestadas tienen menos de 55 libras en su casa?

- ¿Realmente se puede decir que una persona de Oaxaca que trabaja gana en promedio \$2,175.24, cuando se vio que más del 70% de esa población gana menos de \$2,175.24?

La media es una medida muy sensible a la variación de los datos. Es decir, si hay un valor que se aleja mucho del resto de los valores, la media se moverá en gran medida hacia ese valor. Por ello se dice que la media no es una medida robusta.

5.4.2. La moda.

La moda es el valor de la variable que más veces se repite, es decir, el valor al que le corresponde una mayor frecuencia. En el caso de datos sin agrupar, la moda es el valor que más veces se repite. En el supuesto de que los datos estén agrupados en intervalos, la moda es el punto medio del intervalo de mayor frecuencia. Cabe destacar que pueden existir varias modas.

A continuación, veremos algunos ejemplos.

5.4.2.1 Ejemplo 1.

Se encontrarán las modas para el ejemplo de la sección 6.3.3.1.

Como se trata de datos sin agrupar, la moda es el valor que más veces se repite. Si se observa el Cuadro I.5-14, se puede ver que 10 es el valor que más se repite (339 personas dijeron tener 10 libras en su casa), por lo que la moda es 10 libras. Es sencillo verificar que a este valor es al que le corresponde una mayor frecuencia (del 9.95%).

Como no hay ningún otro valor que se repita 339 veces (o más veces), se dice la moda es única.

5.4.2.2 Ejemplo 2.

Se encontrarán las modas para el ejemplo de la sección 6.3.3.2.

Como se trata de datos agrupados en intervalos, la moda es el punto medio del intervalo con mayor frecuencia. En el Cuadro 1.5-17 se puede observar fácilmente que el primer intervalo es el de mayor frecuencia (el lector puede ver que ese intervalo acumula el 30.68% de los datos). Como el intervalo tiene como límite inferior a 1, y como límite superior a 10, la moda es:

$$Moda = \frac{1+10}{2} = \frac{11}{2} = 5.5.$$

Y se dice que la moda es que una persona tenga 6 libros en su casa. Como no hay ningún otro intervalo que acumule el 30.68% de las frecuencias, se dice que la moda es única.

5.4.3 La mediana.

La mediana es el valor de la variable que deja por encima y por debajo de sí a la mitad de las observaciones.

En el caso de que los datos estén sin agrupar, el procedimiento para el cálculo de la mediana es:

1. Se ordenan los datos de menor a mayor.
2. Si el número de observaciones es impar, el valor central es la mediana.
3. Si el número de observaciones es par, existen dos valores centrales. Entonces, la mediana es el valor medio de esos dos valores.

En el caso de que los datos estén agrupados, la mediana es el percentil 50, lo que también ocurre cuando los datos están sin agrupar, note que el algoritmo para encontrar la mediana encontraría también al percentil 50.

Veamos algunos ejemplos.

5.4.3.1 Ejemplo 1.

Se calculará la mediana para el ejemplo de la sección 5.3.3.1.

El Cuadro I.5-10 ya muestra a los valores ordenados de mayor a menor. Además, como el número de observaciones (el número de delegaciones) es par, la mediana se encontrará en el punto medio de los valores 8 y 9, que son los valores centrales. Entonces:

$$Md = \frac{12159 + 12571}{2} = 12365.$$

Y se dice que el 50% de las delegaciones tiene menos de 12,365 viviendas particulares con techos de materiales ligeros, naturales y precarios, y el 50% de las delegaciones tiene más de 12,365 viviendas particulares con esa propiedad. Por otro lado, la media para esos datos es 16,597 viviendas. Esto quiere decir que la media está por encima de la mediana, y que los valores de las delegaciones Gustavo A. Madero e Iztapalapa influenciaron a la media.

5.4.3.2 Ejemplo 2.

Se calculará la mediana para el ejemplo de la sección 5.3.3.2. Para ello, se utilizará la información del Cuadro I.5-13.

Paso 1. El número de observaciones que representa al 50% de la distribución de frecuencias es: $50\left(\frac{16}{100}\right) = 8$.

Paso 2. C_{50} se encuentra en el intervalo 2.

Paso 3. Aplicando la fórmula:

$$C_{50} = 10000 + \frac{9999}{5}(8 - 6) = 10000 + (1999.8)(2) = 10000 + 3999.6 = 13999.6.$$

Y se dice que el 50% de las delegaciones tiene menos de 14,000 viviendas particulares con techos de materiales ligeros naturales y precarios, y

que el 50% de las delegaciones tiene más de 14,000 viviendas particulares con esa propiedad. En este caso, la mediana calculada con datos agrupados sobreestimó a la mediana real.

5.4.3.3 Ejemplo 3.

Se calculará la mediana para el ejemplo de la sección 5.3.4.1.

El Cuadro I.5-14 ya nos muestra a los datos ordenados de menor a mayor. Además, como el número de observaciones (el número de personas que respondieron a esa pregunta) es par, la mediana se encontrará en el punto medio de los valores 1,703 y 1,704. En la tabla 16 se puede observar que las observaciones 1,703 y 1,704 corresponden al mismo valor, 20. Por ello, la mediana es 20.

Entonces, se dice que el 50% de la población entrevistada tiene 20 libros o menos, y que el 50% de la población entrevistada tiene 20 libros o más en su casa. La media para estos datos había sido de 55, esto quiere decir que la media se vio fuertemente afectada por las personas que tenían una gran cantidad de libros en su casa.

5.4.3.4 Ejemplo 4.

Se calculará la mediana para el ejemplo de la sección 5.3.4.2. Los datos serán tomados del Cuadro I.5-17.

Paso 1. El número de observaciones que representa al 50% de la distribución de frecuencias es: $50\left(\frac{3406}{100}\right) = 1703..$

Paso 2. C_{50} se encuentra en el intervalo 3.

Paso 3. Aplicando la fórmula:

$$C_{50} = 21 + \frac{9}{363}(1703 - 1698) = 21 + (0.0247933)(3) = 21 + 0.0743799 = 21.07.$$

Y se dice que el 50% de las personas encuestadas tienen 21 libros o menos en su casa, mientras que el 50% de las personas encuestadas tienen 21 libros o más en su casa. En este caso, la mediana calculada para datos agrupados estimó bastante bien a la mediana real, sobreestimándola ligeramente.

5.4.3.5 Ejemplo 5.

Se calculará la mediana para el ejemplo de la sección 5.3.5.1. Los datos para la realización del ejemplo se pueden ver en el Cuadro 1.5-18.

Como la mediana es el percentil 50, también es el decil 5, y el cuartil 2. En el ejemplo de la sección 6.3.4.1 se calculó al cuartil 2, y se vio que el valor era de \$1,284.51. Entonces, la mediana coincide con ese valor. La media fue de \$2,175.24. Esto quiere decir que la media se vio afectada por las personas de Oaxaca que tenían un ingreso muy grande por su trabajo.

5.4.3.6 Conclusión.

Se ha visto que la mediana tiene la propiedad de robustez. Es decir, la mediana no se ve afectada por valores muy grandes o valores muy pequeños dentro de los datos. Si bien se podría concluir que la mediana es la medida de tendencia central que se debe utilizar en la mayor parte de los casos, no se debe llegar a esa conclusión. La media es la estadística de tendencia central mayormente utilizada debido a su sencillez para el cálculo (es mucho más sencilla de calcular que la mediana, principalmente cuando se tiene una gran cantidad de datos). Además, como se verá en capítulos posteriores, la media goza de muchas propiedades estadísticas de las que la mediana carece. Es aconsejable utilizar la mediana únicamente cuando se sepa o se sospeche que la variación de los datos es muy grande. En la siguiente sección se verán las medidas de dispersión que nos permitirán decidir sobre cuándo vale la pena utilizar a la mediana en lugar de a la media.

5.5 Medidas de dispersión.

Cuando tenemos datos referentes a alguna variable, en muchos casos, resulta importante preguntarnos qué tan grande es la diferencia entre el valor mayor y el menor, o bien, qué tan grande es la diferencia entre los valores extremos y los puntos medios. También resulta importante saber si existen 'lugares' donde se agrupan la mayor parte de los valores. Por ejemplo, si consideramos a los obreros de una fábrica, nos gustaría saber si todos ganan lo mismo, o bien, si hay obreros que ganen más dinero que otros. En este caso, también esperaríamos que si existen diferencias en el salario, éstas no fueran muy grandes. Pero, si consideramos además de a los obreros, a los empleados de confianza, y a lo patrones, entonces esperaríamos que hubieran diferencias importantes en los salarios. Para todo este tipo de situaciones, se utilizan las medidas de dispersión, que, como su nombre lo indica, nos pueden decir si existen grandes diferencias en los valores (es decir, una gran dispersión), o bien, si las diferencias son pequeñas, todo en base a fórmulas que nos permiten realizar el análisis de forma objetiva, y sin tener que analizar a todos los valores.

5.5.1 La amplitud.

También conocida como el recorrido, o el rango total, la amplitud es la más sencilla (en su cálculo) de las medidas de dispersión. Su valor es la diferencia entre el valor máximo y el valor mínimo, para datos no agrupados. Para datos agrupados, el valor es el límite superior de la categoría mayor, menos el valor del límite inferior de la categoría menor. En ambos casos, se supone que los datos, o los intervalos, están ordenados de menor a mayor. Veremos a continuación algunos ejemplos.

5.5.1.1 Ejemplo 1

En la Encuesta Nacional de Juventud, realizada por el Instituto Mexicano de la Juventud a nivel nacional en el año 2000, realizada a jóvenes de entre 12 y 29 años de edad, se preguntó ¿Cuántos hijos quisieras tener en

tu vida? Esa pregunta se aplicó únicamente a jóvenes entre 15 y 29 años de edad, y el cuadro I.5-22 muestra las frecuencias obtenidas. Se obtendrá la amplitud.

CUADRO I.5-22
¿Cuántos hijos quisieras tener en tu vida?*

Número de hijos	Frecuencia
1	1852
2	16473
3	11951
4	3711
5	853
6	314
7	74
8	53
9	39
10	24
11	40
12	24
13	13
14	6
15	4
17	2
18	1
19	1
20	7
21	7
22	11
23	1
24	1
25	1
29	1
30	3
31	2
32	1
33	2
40	1
41	2
42	3
62	1
63	1
80	1

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: *Encuesta Nacional de Juventud, 2000.* Instituto Mexicano de la Juventud.

Nuestros datos están ordenados de menor a mayor y son datos sin agrupar. Entonces:

$$Amplitud = X_N - X_1 = 80 - 1 = 79.$$

Esto nos indica que hay una diferencia de 79 hijos entre los jóvenes que desean tener el menor número de hijos, y los jóvenes que desean tener el mayor número de hijos. Hay que destacar que esta medida es muy sensible a los valores extremos, como el ejemplo ilustra.

5.5.1.2 Ejemplo 2.

Si agrupamos a los datos anteriores, en 6 categorías (véase Cuadro I.5-23), la amplitud no se ve afectada.

CUADRO I.5-23
¿Cuántos hijos quisieras tener en tu vida?*

Número de hijos	Frecuencia
1	1852
2	16473
3	11951
4	3711
5	853
De 6 a 80	641

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Juventud, 2000. Instituto Mexicano de la Juventud.

$$Amplitud = \overline{\lim}I_N - \underline{\lim}I_1 = 80 - 1 = 79.$$

En términos generales, si los intervalos son construidos de tal forma que el límite inferior del primer intervalo corresponde al valor mínimo de los datos sin agrupar, y el límite superior del último intervalo corresponde al valor máximo de los datos sin agrupar, la amplitud no se ve afectada por sus cálculos.

Para concluir, se tiene que decir que la amplitud es una medida que además de verse fuertemente afectada por los valores extremos, por su propia construcción, sólo toma en cuenta a dos valores, que podrían haber sido efecto

de errores, o del azar. Por ejemplo, Sólo un joven contestó que le gustaría tener 80 hijos en toda su vida. Si no hubiese sido entrevistado, la amplitud habría pasado de 79 a 62, que es una diferencia importante. La amplitud como medida de dispersión, debe utilizarse únicamente cuando se trata de obtener un cálculo rápido y sin profundizar sobre la naturaleza de los datos. En general es poco socorrido.

5.5.2 La amplitud intercuartil.

También conocida como recorrido intercuartílico, o como rango intercuartil, mejora la medida de la amplitud, al ser más sensible a la propia concentración de los datos. Su cálculo es igualmente simple: Es la diferencia entre el tercer y el primer cuartil (de ahí su nombre). Veamos un ejemplo.

5.5.2.1 Ejemplo

Se calculará la amplitud intercuartil para el ejemplo de la sección 5.5.1.1,

Se deja al lector verificar que $Q_1 = Q_2 = 2$, y $Q_3 = 3$. Entonces, el rango intercuartil, RIQ, es:

$$RIQ = Q_3 - Q_1 = 3 - 2 = 1.$$

Si recordamos que el cuartil 1 es el que deja por debajo al 25% de los datos, mientras que el cuartil 3 es el que deja por encima al 25% de los datos, concluimos que el 25% de los jóvenes entre 15 y 29 años de edad desean tener 2 hijos o menos a lo largo de su vida, mientras que otro 25% de los jóvenes entre 15 y 29 años de edad desean tener 3 o más hijos a lo largo de su vida. En consecuencia, el 50% de los jóvenes entre 15 y 29 años de edad, desean tener entre 2 y 3 hijos a lo largo de su vida. En otras palabras, para el 50% de los jóvenes entre 15 y 29 años, la amplitud es de 1 hijo, es decir, para el 50% de los jóvenes ubicados en los datos centrales de la encuesta, los valores extremos son tener 2 hijos, o tener 3 hijos, y por lo tanto, su diferencia

es de 1 hijo. También, podemos observar que los datos centrales de la encuesta se encuentran muy concentrados, es decir, la mitad de la población encuestada se encuentra entre los dos valores ya mencionados. Esa es la utilidad del rango intercuartil, observar qué tanto difieren los extremos centrales.

En oposición a la amplitud, la amplitud intercuartil no toma en cuenta a los valores extremos, y sólo utiliza al 50% de los datos para su cálculo, mientras que desprecia a los valores menores o a los mayores, cuando esos son precisamente los que pueden llegar a marcar tendencias clave para la interpretación y análisis de la información. Por ejemplo, si preguntáramos: “En una escala de 0 a 10, donde 0 es totalmente en desacuerdo y 10 es totalmente de acuerdo, dígame ¿Qué tan de acuerdo está usted con que a un violador se le aplique la pena de muerte?”. Si suponemos que el cuartil 1 toma el valor de 5, y el cuartil 3 toma el valor de 8, entonces el rango intercuartil sería de 3. En cambio, si suponemos que la amplitud tomó un valor de 10, podríamos saber que los valores extremos fueron utilizados, y con ese dato, podríamos analizar a las características de la población que estuvo totalmente de acuerdo o totalmente en desacuerdo, para ver si existe algún rasgo que los defina. En este ejemplo hipotético, el rango intercuartil no nos habría sido útil para saber si se habían utilizado los valores extremos o no.

5.5.3 La amplitud semiintercuartil.

Es otra medida de dispersión. A esta medida se le utiliza como un parámetro para ver qué tanto se separan los datos del valor de la mediana. Es decir, al calcular la mediana obtenemos el número por debajo del cual está el 50% de la distribución. Igualmente, el 50% de los datos se encuentran por arriba de la mediana. A su vez, entre el cuartil 1 y la mediana, se encuentra el 25% de los datos, e igualmente entre la mediana y el cuartil 3. Si la distancia entre el cuartil 1 y la mediana es igual que la distancia entre la mediana y el cuartil 3, entonces, la mitad del rango intercuartil es igual a la distancia entre la mediana y el cuartil 3, así como la distancia entre la mediana y el cuartil 1. Si la distancia entre el cuartil 1 y la mediana es mayor que la distancia entre la

mediana y el cuartil 3, entonces, la distancia entre la mediana y el cuartil 1 es mayor a la mitad del rango intercuartil, mientras que la distancia entre la mediana y el cuartil 3 es menor a la mitad del rango intercuartil. Si la distancia entre el cuartil 1 y la mediana es menor que la distancia entre la mediana y el cuartil 3, entonces, la distancia entre la mediana y el cuartil 1 es menor a la mitad de rango intercuartil, mientras que la distancia entre la mediana y el cuartil 3 es mayor a la mitad del rango intercuartil. De los hechos mencionados anteriormente, podemos concluir que la mitad del rango intercuartil nos puede decir que tan bien centrados se encuentran los datos, respecto a la mediana, y por ello, la mitad del rango intercuartil se utiliza como medida de dispersión respecto a la mediana, y se le llama la amplitud semiintercuartil, o comúnmente, rango semiintercuartil. Veamos algunos ejemplos.

5.5.3.1 Ejemplo 1.

Se calculará el rango semiintercuartil para el ejemplo de la sección 5.5.2.1. Como $RIQ = 1$, entonces, el rango semiintercuartil, RSQ , es

$$RSQ = \frac{Q_3 - Q_1}{2} = \frac{1}{2}.$$

La mediana toma el valor de 2 (se dejan al lector los cálculos), y es igual al cuartil 1, mientras que el cuartil 3 toma el valor de 3. Como la distancia entre la mediana y el cuartil 3 (1) es mayor a la distancia entre la mediana y el cuartil 1 (0), la distancia entre la mediana y el cuartil 3 es mayor que RSQ , denotando que los datos se encuentran más concentrados en el cuartil 1.

5.5.3.2 Ejemplo 2.

Se tomarán las respuestas a la pregunta ¿Qué edad tenías cuando nació tu primer hijo? De la Encuesta Nacional de Juventud (véase cuadro I.5-24). Esta pregunta fue aplicada a jóvenes entre 15 y 29 años de edad que hubieran tenido hijos. Con estos datos se calculará el rango semiintercuartil.

CUADRO I.5-24

¿Qué edad tenías cuando nació tu primer hijo(a)?*

Edad	Frecuencia	% Frecuencia Acumulada
10	9	0.08%
11	1	0.09%
12	6	0.14%
13	34	0.43%
14	97	1.27%
15	347	4.25%
16	734	10.57%
17	1169	20.63%
18	1516	33.68%
19	1619	47.61%
20	1492	60.45%
21	1288	71.54%
22	975	79.93%
23	816	86.95%
24	562	91.79%
25	392	95.16%
26	297	97.72%
27	171	99.19%
28	75	99.84%
29	19	100.00%

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Juventud, 2000. Instituto Mexicano de la Juventud.

Resulta claro que $Q_1 = 17$, $Q_2 = 20$, y $Q_3 = 22$. Entonces,

$$RSQ = \frac{Q_3 - Q_1}{2} = \frac{22 - 17}{2} = 2.5.$$

Observemos que entre los 17 y los 23 años de edad ($Q_2 \pm ISQ$), es decir la distancia de la mediana al rango semiintercuartil, más del 50% de los jóvenes entrevistados tuvieron a su primer hijo.

5.5.3.3 Ejemplo 3.

Ahora, se calculará el rango semiintercuartil para las respuestas a la pregunta “En el último mes, ¿Cuántas veces fuiste a fiestas?” (véase Cuadro I.5-25), tomada de la Encuesta Nacional de Juventud, y aplicada a jóvenes entre 12 y 29 años de edad.

CUADRO I.5-25
En el último mes, ¿Cuántas veces fuiste a fiestas?*

Número de Veces	Frecuencia	% Frecuencia Acumulada
1	10380	39.18%
2	8080	69.68%
3	3963	84.64%
4	2079	92.48%
5	894	95.86%
6	413	97.42%
7	139	97.94%
8	187	98.65%
9	28	98.75%
10	202	99.52%
11	4	99.53%
12	26	99.63%
13	5	99.65%
15	30	99.76%
16	2	99.77%
17	1	99.77%
18	1	99.78%
20	44	99.94%
22	2	99.95%
25	1	99.95%
26	1	99.96%
30	11	100.00%

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Juventud, 2000. Instituto Mexicano de la Juventud.

Resulta claro que $Q_1 = 1$, $Q_2 = 2$, y $Q_3 = 3$. Entonces, $RSQ = \frac{Q_3 - Q_1}{2} = \frac{3-1}{2} = 1$. Observemos que los jóvenes que fueron a entre una y tres fiestas ($Q_2 \pm ISQ$) en el último mes representan más del 50% de los jóvenes entrevistados. También se puede destacar que en este caso, los valores se encuentran bien centrados. La mediana se encuentra en el punto medio de los cuartiles 1 y 3, y, por lo tanto, la distancia de la mediana a cualquiera de los otros dos cuartiles es igual al valor de RSQ.

5.5.3.4 Conclusión.

El rango semiintercuartil es una medida que nos permite medir la dispersión de los datos respecto a la mediana. Su valor nos indica que $RSQ \pm$ unidades a la redonda de la mediana se encuentra la mayor parte de los datos.

Cuando utilizamos a la mediana como medida de tendencia central, el rango semiintercuartil es el ideal para medir la dispersión.

5.5.4 La desviación media.

Las medidas de dispersión tratadas hasta ahora, sólo utilizan algunos valores. La amplitud únicamente utiliza a los dos valores extremos; la amplitud intercuartil y la amplitud semiintercuartil, utilizan sólo a los datos que se encuentran entre los cuartiles 1 y 3. Una medida de dispersión que sí contempla a todos los datos es la desviación media. La desviación media (o desviación típica) es el promedio de las distancias de un conjunto de puntuaciones respecto a la media. A continuación se presentan las fórmulas para datos sin agrupar, y para datos agrupados, seguidos de algunos ejemplos.

5.5.4.1 Fórmula para datos sin agrupar.

Si seguimos la definición, la desviación media es el promedio de las distancias de un conjunto de puntuaciones respecto a la media, es decir:

$$\frac{d(X_1, \bar{X}) + d(X_2, \bar{X}) + \dots + d(X_n, \bar{X})}{n}, \text{ donde,}$$

$d(X_i, \bar{X})$ es la distancia entre el valor de la unidad i -ésima y el valor medio.

n son el número de datos que tenemos.

Como $d(X_i, \bar{X}) \geq 0$ para todo valor, y se trabaja en una dimensión, $d(X_i, \bar{X})$ es la diferencia del valor mayor y el valor menor. Es decir, la desviación media es la suma de los valores positivos de las diferencias entre los valores y su media. Por ello, la fórmula para datos sin agrupar utiliza a los valores absolutos, y es:

$$DM = \frac{|X_1 - \bar{X}| + |X_2 - \bar{X}| + \dots + |X_n - \bar{X}|}{n}.$$

5.5.4.2 Fórmula para datos agrupados.

La fórmula para el cálculo de la desviación media para datos agrupados. es una derivación de la anterior, y es la siguiente:

$$DM = \frac{n_1 |X_1 - \bar{X}| + n_2 |X_2 - \bar{X}| + \dots + n_N |X_n - \bar{X}|}{n_1 + n_2 + \dots + n_N}, \text{ donde}$$

n_i es la frecuencia absoluta del intervalo i -ésimo.

X_i es el punto medio del intervalo i -ésimo.

\bar{X} es la media calculada para datos agrupados.

5.5.4.3 Ejemplo 1.

Se calculará la desviación típica para el ejemplo de la sección 5.3.3.1. Para ello, se encontró la media en la sección 5.4.1.1, y tomó un valor de 16,596.56. Tendremos que calcular las diferencias entre los valores y esa media, para luego, tomar su valor absoluto, y finalmente sumarmas y promediarlas (véase Cuadro I.5-26).

CUADRO I.5-26
Desviación media de las viviendas con materiales precarios en las delegaciones del Distrito Federal*

	Delegación	Viv c/mat prec**	Media	X_i -media	Valor absoluto
1	Benito Juárez	2,384	16,596.56	-14,213	14212.56
2	Cuauhtémoc	4,194	16,596.56	-12,403	12402.56
3	Milpa Alta	6,254	16,596.56	-10,343	10342.56
4	Cuajimalpa de Morelos	7,241	16,596.56	-9,356	9355.56
5	Miguel Hidalgo	8,127	16,596.56	-8,470	8469.56
6	Venustiano Carranza	8,728	16,596.56	-7,869	7868.56
7	Tláhuac	11,905	16,596.56	-4,692	4691.56
8	Coyoacán	12,159	16,596.56	-4,438	4437.56
9	La Magdalena Contreras	12,571	16,596.56	-4,026	4025.56
10	Azcapotzalco	13,172	16,596.56	-3,425	3424.56
11	Iztacalco	13,358	16,596.56	-3,239	3238.56
12	Xochimilco	20,008	16,596.56	3,411	3411.44
13	Álvaro Obregón	21,203	16,596.56	4,606	4606.44

14	Tlalpan	24,838	16,596.56	8,241	8241.44
15	Gustavo A. Madero	34,570	16,596.56	17,973	17973.44
16	Iztapalapa	64,833	16,596.56	48,236	48236.44
	SUMAS	265,545	265544.96	0	164938.36

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE 2000, INEGI.

**Viviendas particulares con techos de materiales ligeros, naturales y precarios.

$$\text{Por lo que, } DM = \frac{164,938.36}{16} = 10,308.65.$$

Se puede apreciar que 11 de las 16 delegaciones tienen entre 6,288 y 26,905 ($\bar{X} \pm DM$) viviendas con techos de materiales ligeros, naturales y precarios, que representan a más del 50%.

También se puede apreciar que el dato que más se aleja de la media es el correspondiente a la delegación Iztapalapa, donde se presenta el valor absoluto mayor.

5.5.4.4 Ejemplo 2.

Se calculará la desviación típica para las respuestas a la pregunta “¿A qué edad tuviste tu primera relación sexual?” (véase Cuadro I.5-27), preguntada en la Encuesta Nacional de Juventud, a los jóvenes entre 15 y 29 años de edad que han tenido relaciones sexuales.

CUADRO I.5-27
¿A qué edad tuviste tu primera relación sexual?

Edad	Frecuencia	Media	X_i -media	Valor absoluto
6	3	17.7203	-35.1609	35.1609
7	3	17.7203	-32.1609	32.1609
8	15	17.7203	-145.8045	145.8045
9	17	17.7203	-148.2451	148.2451
10	18	17.7203	-138.9654	138.9654
11	70	17.7203	-470.4210	470.4210
12	172	17.7203	-983.8916	983.8916
13	391	17.7203	-1,845.6373	1,845.6373
14	1,073	17.7203	-3,991.8819	3,991.8819

15	2,168	17.7203	-5,897.6104	5,897.6104
16	2,732	17.7203	-4,699.8596	4,699.8596
17	3,113	17.7203	-2,242.2939	2,242.2939
18	3,260	17.7203	911.8220	911.8220
19	1,893	17.7203	2,422.4721	2,422.4721
20	1,645	17.7203	3,750.1065	3,750.1065
21	850	17.7203	2,787.7450	2,787.7450
22	690	17.7203	2,952.9930	2,952.9930
23	453	17.7203	2,391.7041	2,391.7041
24	311	17.7203	1,952.9867	1,952.9867
25	230	17.7203	1,674.3310	1,674.3310
26	121	17.7203	1,001.8437	1,001.8437
27	60	17.7203	556.7820	556.7820
28	19	17.7203	195.3143	195.3143
29	3	17.7203	33.8391	33.8391
SUMAS	19,310	425.2872	0.0000	41,263.8720

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Juventud, 2000. Instituto Mexicano de la Juventud.

En primer lugar se tiene que destacar que los datos presentados en el Cuadro I.5-27 son datos sin agrupar presentados de forma resumida. La primera columna muestra la edad, la segunda columna muestra la cantidad de jóvenes que contestaron esa edad, la tercera columna muestra a la media, la cuarta columna muestra a la diferencia entre el valor y la media, multiplicado por su frecuencia, y la sexta columna muestra a los valores absolutos de la quinta columna. Utilizando la fórmula para el cálculo de la desviación media, obtenemos que:

$$DM = \frac{41,263.8720}{19,310} = 2.1369.$$

Se puede observar que más del 50% de los jóvenes que contestaron a esta pregunta tuvieron su primera relación sexual entre los 16 y los 20 años ($\bar{X} \pm DM$), y que las edades más alejadas de esta media, son los 6 años y los 29 años de edad.

5.5.4.5 Ejemplo 3.

Se calculará la desviación típica para el ejemplo de la sección 5.3.5.1. Como se trata de datos agrupados, para el cálculo de la desviación media, se deben calcular los valores medios de los intervalos, los valores absolutos de las diferencias respecto a la media (la media toma el valor de 2175.2404, y fue obtenida en la sección 5.4.1.5), y la suma de las frecuencias (véase Cuadro I.5-28).

CUADRO I.5-28
Dispersión respecto a la media del nivel de ingreso de las personas en Oaxaca*

Cifras en salarios mínimos	n_i	X_i	Media	$n_i(X_i - \text{media})$	Valor Absoluto
0 - 1/4 SM	9,047	\$122.63	2,175.2404	-18,570,011.5238	18,570,011.5238
1/4 - 1/2 SM	10,444	\$367.88	2,175.2404	-18,876,124.2376	18,876,124.2376
1/2 - 1 SM	26,178	\$735.75	2,175.2404	-37,682,979.6912	37,682,979.6912
1 - 2 SM	36,452	\$1,471.50	2,175.2404	-25,652,745.0608	25,652,745.0608
2 - 3 SM	13,331	\$2,452.50	2,175.2404	3,696,147.7276	3,696,147.7276
3 - 4 SM	7,790	\$3,433.50	2,175.2404	9,801,842.2840	9,801,842.2840
4 - 5 SM	5,008	\$4,414.50	2,175.2404	11,214,212.0768	11,214,212.0768
5 - 6 SM	1,734	\$5,395.50	2,175.2404	5,583,930.1464	5,583,930.1464
6 - 7 SM	1,303	\$6,376.50	2,175.2404	5,474,241.2588	5,474,241.2588
7 - 8 SM	410	\$7,357.50	2,175.2404	2,124,726.4360	2,124,726.4360
8 - 9 SM	697	\$8,338.50	2,175.2404	4,295,791.9412	4,295,791.9412
9 - 10 SM	141	\$9,319.50	2,175.2404	1,007,340.6036	1,007,340.6036
10 - 15 SM	746	\$12,267.50	2,175.2404	7,528,825.6616	7,528,825.6616
15 - 20 SM	238	\$17,167.50	2,175.2404	3,568,157.7848	3,568,157.7848
20 - 30 SM	188	\$24,525.00	2,175.2404	4,201,754.8048	4,201,754.8048
30 - 50 SM	99	\$39,240.00	2,175.2404	3,669,411.2004	3,669,411.2004
50 - 100 SM	20	\$73,575.00	2,175.2404	1,427,995.1920	1,427,995.1920
100SM o más	68	\$549,050.00	2,175.2404	37,187,483.6528	37,187,483.6528
SUMAS	113,894	765,610.75	39,154.3272	0.0000	201,563,721.2842

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: XII Censo General de Población y Vivienda. Cuestionario Ampliado. INEGI, 2000.

Con esos datos, se obtiene la desviación media, que es:

$$DM = \frac{201,563,721.2842}{113,894} = 1,769.7884 = \$1,769.79.$$

Se puede apreciar que más del 50% de las personas que contestaron a esta encuesta, ganaban por su trabajo más de \$405.25 y menos de \$3,945.06 ($\bar{X} \pm DM$).

5.5.4.6 Conclusiones

La desviación media es una medida de dispersión que, si se utiliza de forma conjunta con la media, nos puede decir en donde se localizan la mayor parte de los valores. Además, tiene la virtud de tomar en cuenta a las diferencias reales de los datos. Por otro lado, en términos generales, se puede decir que entre más grande sea la desviación media, mayor será la variación de los datos.

No obstante lo anterior, resulta inconveniente por varias cosas. Debido a que utiliza para su cálculo a los valores absolutos, es difícil de manipular matemáticamente. Su interpretación teórica resulta complicada, como veremos en secciones subsecuentes, y no es la más usada en la literatura (por estas razones). A la desviación media se le debe utilizar sólo de forma descriptiva, y no se debe esperar más de ella. La medida de dispersión que veremos a continuación es la más utilizada de todas, y goza de muchas propiedades estadísticas que se analizarán en capítulos posteriores.

5.5.5 La desviación estándar.

La desviación estándar es una medida similar a la desviación media, en el sentido de que se basa en las diferencias existentes entre la media y cada valor. En el caso de la desviación media, se tomó el valor absoluto de las diferencias, para que todas fueran positivas, y así, no se anularan unas con otras. En el caso de la desviación estándar, el procedimiento que se utiliza para obtener valores positivos de las diferencias es elevar al cuadrado las diferencias. Al utilizar ese procedimiento, se estarían utilizando valores

cuadrados (hijos al cuadrado, dinero al cuadrado, etc), por ello, al valor obtenido de la suma se le aplica la raíz cuadrada, para regresar a los datos a su estructura original.

La fórmula para datos no agrupados es:

$$s = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n}}, \text{ donde,}$$

s es la desviación estándar.

n es el número de observaciones.

X_i es el valor de la variable, para el caso i -ésimo.

\bar{X} es la media.

Para datos agrupados, la fórmula es la siguiente:

$$s = \sqrt{\frac{(X_1 - \bar{X})^2 n_1 + (X_2 - \bar{X})^2 n_2 + \dots + (X_N - \bar{X})^2 n_N}{n_1 + n_2 + \dots + n_N}}, \text{ donde,}$$

s es la desviación estándar.

n_i es la frecuencia absoluta del intervalo i -ésimo.

X_i es el valor medio del intervalo i -ésimo.

\bar{X} es la media, calculada para datos agrupados.

A continuación se presentan algunos ejemplos.

5.5.5.1 Ejemplo 1

Se calcularán la media y la desviación estándar para el ejemplo de la sección 5.2.2. El cálculo de la media se realiza sumando las poblaciones de los cinco municipios, y dividiéndola entre cinco (el número de municipios). Entonces:

$$\bar{X} = \frac{11,812 + 45,989 + 63,864 + 105,469 + 196,907}{5} = \frac{424,041}{5} = 84,808.20 .$$

Y se dice que en promedio, cada municipio de Baja California Sur tiene 84,808 habitantes. Ahora calcularemos la desviación estándar. Para ello, y por tratarse de datos no agrupados, se necesita calcular, además de la media, los cuadrados de las diferencias (véase cuadro I.5-29). Con esos valores, la aplicación de la fórmula resulta sencilla, y es:

$$s = \sqrt{\frac{20,267,044,634.80}{5}} = \sqrt{4,053,408,926.96} = 63,666.39.$$

CUADRO I.5-29
Desviación estándar de la población de Baja California Sur por municipio*

Municipio	Población Total	Media	Xi - media	(Xi - media)^2
Loreto	11,812	84,808.20	-72,996.20	5,328,445,214.44
Mulegé	45,989	84,808.20	-38,819.20	1,506,930,288.64
Comondú	63,864	84,808.20	-20,944.20	438,659,513.64
Los Cabos	105,469	84,808.20	20,660.80	426,868,656.64
La Paz	196,907	84,808.20	112,098.80	12,566,140,961.44
SUMAS	424,041	424,041.00	0.00	20,267,044,634.80

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE 2000, INEGI.

Hay que destacar que tres de los cinco municipios (el 60%) tiene entre 21,140 y 148,475 ($\bar{X} \pm s$) habitantes. Además, como el valor de s es grande, se puede concluir que existe una dispersión grande entre las poblaciones de los municipios, lo cual se corrobora al observar el Cuadro I.5-29.

5.5.5.2 Ejemplo 2

Se calculará la desviación estándar para el ejemplo de la sección 5.3.3.1. En este caso, la media fue calculada en la sección 5.4.1.1 y tuvo un valor de 16,596.56. Adicionalmente se necesita calcular el cuadrado de las diferencias (véase Cuadro I.5-30). Una vez obtenidos los datos, se puede aplicar la fórmula.

CUADRO I.5-30
Desviación estándar de las viviendas con materiales precarios en las delegaciones del Distrito Federal*

	Delegación	Viv c/mat prec**	Media	$X_i - \text{media}$	$(X_i - \text{media})^2$
1	Benito Juárez	2,384	16,596.56	-14,213.00	202,009,369.00
2	Cuauhtémoc	4,194	16,596.56	-12,403.00	153,834,409.00
3	Milpa Alta	6,254	16,596.56	-10,343.00	106,977,649.00
4	Cuajimalpa de Morelos	7,241	16,596.56	-9,356.00	87,534,736.00
5	Miguel Hidalgo	8,127	16,596.56	-8,470.00	71,740,900.00
6	Venustiano Carranza	8,728	16,596.56	-7,869.00	61,921,161.00
7	Tláhuac	11,905	16,596.56	-4,692.00	22,014,864.00
8	Coyoacán	12,159	16,596.56	-4,438.00	19,695,844.00
9	La Magdalena Contreras	12,571	16,596.56	-4,026.00	16,208,676.00
10	Azcapotzalco	13,172	16,596.56	-3,425.00	11,730,625.00
11	Iztacalco	13,358	16,596.56	-3,239.00	10,491,121.00
12	Xochimilco	20,008	16,596.56	3,411.00	11,634,921.00
13	Álvaro Obregón	21,203	16,596.56	4,606.00	21,215,236.00
14	Tlalpan	24,838	16,596.56	8,241.00	67,914,081.00
15	Gustavo A. Madero	34,570	16,596.56	17,973.00	323,028,729.00
16	Iztapalapa	64,833	16,596.56	48,236.00	2,326,711,696.00
	SUMAS	265,545	265,545	0.00	3,514,664,017.00

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE 2000, INEGI.

**Viviendas particulares con techos de materiales ligeros, naturales y precarios.

$$s = \sqrt{\frac{3,514,664,017}{16}} = \sqrt{219,666,501.0625} = 14,821.15 .$$

En la sección 6.5.4.3 se obtuvo la desviación media para estos datos, y tuvo un valor de 10,308.65. Es decir, la desviación media fue menor que la desviación estándar. Lo anterior se debe a que la desviación media no se vio tan afectada como la desviación estándar por los valores extremos (en este caso, el de la delegación Iztapalapa, que, al elevar al cuadrado su diferencia respecto a la media, aportó más del 50% del valor de la desviación estándar). Es decir, la desviación estándar es más sensible a valores muy grandes o muy pequeños (respecto a la media). Resulta claro que en $(\bar{X} \pm s)$ están contempladas más del 50% de las delegaciones.

5.5.5.3 Ejemplo 3

Se calculará la desviación estándar para el ejemplo de la sección 5.5.3.2. Para obtener este resultado es necesario primero calcular la media. Para ello, se calculan las sumas de los productos de la edad por la frecuencia, para ser dividido entre el total de las frecuencias (véase Cuadro I.5-31).

Primero se calcula la media, que es $\bar{X} = \frac{232,228}{11,619} = 19.99$. Una vez

calculada la media, se deben calcular los cuadrados de las diferencias, ponderados por las frecuencias (véase Cuadro I.5-31). Con esos cálculos, se puede aplicar la fórmula para el cálculo de la desviación estándar.

CUADRO I.5-31
¿Qué edad tenías cuando nació tu primer hijo(a)?*

¿Qué edad tenías cuando nació tu primer hijo(a)?					
Edad	n_i	$X_i n_i$	Media	$(X_i - \text{Media})$	$n_i(X_i - \text{Media})^2$
10	9	90	19.99	-89.88	897.65
11	1	11	19.99	-8.99	80.76
12	6	72	19.99	-47.92	382.75
13	34	442	19.99	-237.56	1,659.78
14	97	1,358	19.99	-580.73	3,476.79
15	347	5,205	19.99	-1,730.46	8,629.66
16	734	11,744	19.99	-2,926.40	11,667.31
17	1169	19,873	19.99	-3,491.71	10,429.44
18	1516	27,288	19.99	-3,012.17	5,984.93
19	1619	30,761	19.99	-1,597.82	1,576.92
20	1492	29,840	19.99	19.52	0.26
21	1288	27,048	19.99	1,304.85	1,321.92
22	975	21,450	19.99	1,962.75	3,951.19
23	816	18,768	19.99	2,458.67	7,408.19
24	562	13,488	19.99	2,255.35	9,050.91
25	392	9,800	19.99	1,965.13	9,851.35
26	297	7,722	19.99	1,785.89	10,738.68
27	171	4,617	19.99	1,199.24	8,410.35
28	75	2,100	19.99	600.98	4,815.71
29	19	551	19.99	171.25	1,543.48
SUMAS	11619	232228	399.74	0.00	101,878.01

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Juventud, 2000. Instituto Mexicano de la Juventud.

$$s = \sqrt{\frac{101,878.01}{11,619}} = \sqrt{8.768225} = 2.96.$$

Resulta claro que más del 50% de los jóvenes que contestaron esta pregunta tuvieron a su primer hijo cuando tenían entre 17 y 23 años de edad ($\bar{X} \pm s$). Para este ejemplo se había calculado el rango semiintercuartil, que tomó un valor de 2.5, es decir, fue menor que la desviación estándar. Esto se debe principalmente a que el rango semiintercuartil toma en cuenta únicamente al 50% de los datos centrales, mientras que la desviación estándar toma en cuenta a todos los datos, incluso a los extremos. De hecho, entre más alejado esté un valor de la media, ese valor incrementará a la desviación estándar en una medida mucho mayor que la desviación media, o que el *RSQ*, éste último, de hecho, no toma en cuenta a los valores muy alejados de la media.

5.5.5.4 Ejemplo 4

Se calculará la desviación estándar para el ejemplo de la sección 5.3.4.1. Para ello, se calculó la media en la sección 5.4.1.5, y tomó un valor de 2175.2404. Además de ese valor, es necesario calcular la suma del cuadrado de las diferencias (véase Cuadro I.5-32).

CUADRO I.5-28
Dispersión respecto a la media del nivel de ingreso de las personas en Oaxaca*

Cifras en salarios mínimos	n_i	X_i	Media	$n_i(X_i - \text{media})$	$n_i(X_i - \text{media})^2$
0 - 1/4 SM	9,047	\$122.63	2,175.24	-18,570,011.52	38,116,905,932.04
1/4 - 1/2 SM	10,444	\$367.88	2,175.24	-18,876,124.24	34,115,865,072.16
1/2 - 1 SM	26,178	\$735.75	2,175.24	-37,682,979.69	54,244,287,508.88
1 - 2 SM	36,452	\$1,471.50	2,175.24	-25,652,745.06	18,052,873,070.19
2 - 3 SM	13,331	\$2,452.50	2,175.24	3,696,147.73	1,024,792,440.50
3 - 4 SM	7,790	\$3,433.50	2,175.24	9,801,842.28	12,333,262,151.53
4 - 5 SM	5,008	\$4,414.50	2,175.24	11,214,212.08	25,111,532,049.41
5 - 6 SM	1,734	\$5,395.50	2,175.24	5,583,930.15	17,981,704,659.67
6 - 7 SM	1,303	\$6,376.50	2,175.24	5,474,241.26	22,998,708,641.25
7 - 8 SM	410	\$7,357.50	2,175.24	2,124,726.44	11,010,883,970.33

8 - 9 SM	697	\$8,338.50	2,175.24	4,295,791.94	26,476,080,921.20
9 - 10 SM	141	\$9,319.50	2,175.24	1,007,340.60	7,196,702,777.74
10 - 15 SM	746	\$12,267.50	2,175.24	7,528,825.66	75,982,863,060.01
15 - 20 SM	238	\$17,167.50	2,175.24	3,568,157.78	53,494,747,803.48
20 - 30 SM	188	\$24,525.00	2,175.24	4,201,754.80	93,908,209,785.42
30 - 50 SM	99	\$39,240.00	2,175.24	3,669,411.20	136,005,844,016.37
50 - 100 SM	20	\$73,575.00	2,175.24	1,427,995.19	101,958,513,418.76
100SM o más	68	\$549,050.00	2,175.24	37,187,483.65	20,336,896,182,753.90
SUMAS	113,894	765,610.75	39,154.33	0.00	21,066,909,960,032.90

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: XII Censo General de Población y Vivienda. Cuestionario Ampliado. INEGI, 2000.

Una vez que se tienen esos valores, se aplica la fórmula para el cálculo de la desviación estándar, y se obtiene:

$$s = \sqrt{\frac{21,066,909,960,032.90}{113,894}} = \sqrt{184,969,444.92} = 13600.35.$$

Salta a la vista en forma inmediata que es un valor muy grande. De hecho, más del 99% de la población encuestada gana entre \$0 y \$15,775.59 ($\bar{X} \pm s$). Este caso muy particular se da porque el último intervalo se aleja tanto del promedio, que influye drásticamente a la desviación estándar. En el Cuadro I.5-32 se puede apreciar que ese intervalo aporta más del 95% del valor de la suma, lo cual afecta severamente a la desviación estándar. De hecho, si no se tomará en cuenta, el valor de la desviación estándar habría sido de 2,535 (se dejan al lector los cálculos), lo que significa que el valor de la diferencia cuadrada ponderada por la frecuencia del último intervalo aportó el 81.36% de la desviación estándar. Ya cuando se calculó la desviación media para este ejemplo se había apreciado que el último intervalo provocaba un crecimiento en el valor de la desviación, sin embargo, no se vio tan afectada (su valor fue de 1,769.79). Es decir, la desviación estándar nos muestra con mucha mayor claridad cuando los datos tienen una gran dispersión.

5.5.5.5 Ejemplo 5

Se calculará la desviación estándar para el ejemplo de la sección 5.5.4.3. Para ello, se debe calcular la suma del cuadrado de las diferencias (véase Cuadro I.5-33). La media fue calculada en esa misma sección y fue de 17.7203.

CUADRO I.5-33
¿A qué edad tuviste tu primera relación sexual? (Desviación estándar)

Edad	Frecuencia	Media	$X_i - \text{media}$	$n_i(X_i - \text{Media})^2$
6	3	17.7203	-35.16	412.10
7	3	17.7203	-32.16	344.77
8	15	17.7203	-145.80	1,417.26
9	17	17.7203	-148.25	1,292.74
10	18	17.7203	-138.97	1,072.85
11	70	17.7203	-470.42	3,161.37
12	172	17.7203	-983.89	5,628.16
13	391	17.7203	-1,845.64	8,711.96
14	1,073	17.7203	-3,991.88	14,851.00
15	2,168	17.7203	-5,897.61	16,043.27
16	2,732	17.7203	-4,699.86	8,085.17
17	3,113	17.7203	-2,242.29	1,615.12
18	3,260	17.7203	911.82	255.04
19	1,893	17.7203	2,422.47	3,100.04
20	1,645	17.7203	3,750.11	8,549.12
21	850	17.7203	2,787.75	9,142.97
22	690	17.7203	2,952.99	12,637.92
23	453	17.7203	2,391.70	12,627.48
24	311	17.7203	1,952.99	12,264.17
25	230	17.7203	1,674.33	12,188.63
26	121	17.7203	1,001.84	8,294.97
27	60	17.7203	556.78	5,166.77
28	19	17.7203	195.31	2,007.77
29	3	17.7203	33.84	381.69
SUMAS	19,310	425.2872	0.00	149,252.34

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Juventud, 2000. Instituto Mexicano de la Juventud.

Ya con esos datos, se aplica la fórmula para el cálculo de la desviación estándar, y se obtiene:

$$s = \sqrt{\frac{149,252.34}{19,310}} = \sqrt{7.729277} = 2.78.$$

Resulta claro que más del 50% de los jóvenes encuestados tuvieron su primera relación sexual entre los 15 y los 20 años de edad ($\bar{X} \pm s$). Por otro lado, la desviación media para este caso fue de 2.14, es decir, fue menor que la desviación estándar. Nuevamente esto se debe a que la desviación estándar se ve más afectada que la desviación media por los datos que se alejan mucho de la media.

5.5.6 La varianza.

La varianza se define como el cuadrado de la desviación estándar, y se denota por s^2 . Estadísticamente, la varianza posee mayor valor teórico que la desviación estándar, y en capítulos posteriores se hará mayor uso de ella, pero por el momento el concepto de desviación estándar es suficiente. Aunque pareciera más práctico calcular la varianza en lugar de la desviación estándar (implica el ahorro del cálculo de una raíz cuadrada), eso sería un error, ya que la varianza presenta datos al cuadrado. Por ejemplo, cuando trabajamos con edades, la desviación estándar también obtiene edades (2 años de edad, 3 años de edad, etc), pero la varianza trabaja con edades al cuadrado, cuya interpretación es impráctica y más difícil.

5.5.7 El coeficiente de variación.

A veces, es conveniente comparar varios grupos en relación a su dispersión. Como ya se ha visto, la desviación estándar puede variar mucho entre diferentes ejemplos. Una operación sencilla que nos permite comparar la dispersión de los grupos es el coeficiente de variación, designado por V . Su fórmula es la siguiente:

$$V = \frac{s}{\bar{X}}$$

A continuación se presentan algunos ejemplos.

5.5.7.1 Ejemplo 1.

Se desean comparar los datos de la sección 5.5.3.2 y 5.5.4.3, es decir, se desea comparar la edad a la que los jóvenes encuestados tuvieron su primera relación sexual con la edad a la que los jóvenes encuestados tuvieron a su primer hijo.

En promedio, los jóvenes entrevistados tuvieron su primera relación sexual a los 17.72 años de edad (véase sección 5.5.4.3), con una desviación estándar de 2.78 (véase sección 5.5.5.5); mientras que tuvieron su primer hijo a los 19.99 años de edad, con una desviación estándar de 2.96 (véase sección 6.5.5.3). Aunque en principio las desviaciones estándar son similares y las medias se alejan por alrededor de 2 años de edad, antes de realizar alguna comparación, se calcularán los coeficientes de variación para ambos casos.

El coeficiente de variación para los jóvenes que contestaron a qué edad habían tenido su primera relación sexual es: $V = \frac{2.78}{17.72} = 0.1569$. A su vez, el coeficiente de variación para los jóvenes que contestaron a qué edad habían tenido su primer hijo es: $V = \frac{2.96}{19.99} = 0.1481$.

De los resultados anteriores se puede apreciar que aunque existe una mayor dispersión entre los jóvenes que contestaron a qué edad habían tenido a su primer hijo que entre los jóvenes que contestaron a qué edad habían tenido su primera relación sexual, su dispersión relativa a la edad promedio se invierte, es decir, hay una menor dispersión entre los jóvenes que contestaron a qué edad habían tenido a su primer hijo, respecto a la edad promedio, que la dispersión entre los jóvenes que contestaron a qué edad habían tenido su

primera relación sexual, respecto a la edad promedio. En conclusión, se puede decir que es más probable que los jóvenes encuestados tengan su primera relación sexual antes o después de la edad promedio, a que tengan a su primer hijo antes o después de la edad promedio.

5.5.7.2 Ejemplo 2

Se desean comparar los datos de la sección 5.5.3.2, es decir, se desea comparar la edad a la que los jóvenes encuestados tuvieron su primera relación sexual, para hombres, y para mujeres.

Para ello, se deben obtener los resultados por separado para hombres y para mujeres (véase cuadro I.5-34). Igualmente se deben calcular las medias, y las desviaciones estándar para cada uno de los casos. Ya con esa información, se procede a calcular el coeficiente de variación.

CUADRO I.5-34
¿A qué edad tuviste tu primera relación sexual? (por sexo)

Edad	Hombre (X)	Mujer (Y)	MediaH**	MediaM***	$X_i - \text{MediaH}$	$Y_i - \text{MediaM}$	$n_i(X_i - \text{MediaH})^2$	$n_i(Y_i - \text{MediaM})^2$
6	2	1	17.21	18.19	-22.42	-12.19	-2,818.97	148.64
7	1	2	17.21	18.19	-10.21	-22.38	-1,064.99	250.51
8	12	3	17.21	18.19	-110.55	-30.58	-9,381.16	311.62
9	12	5	17.21	18.19	-98.55	-45.96	-6,645.74	422.44
10	12	6	17.21	18.19	-86.55	-49.15	-4,501.59	402.63
11	49	21	17.21	18.19	-304.39	-151.03	-11,746.55	1,086.15
12	113	59	17.21	18.19	-588.97	-365.31	-15,999.83	2,261.95
13	261	130	17.21	18.19	-1,099.36	-674.93	-19,504.51	3,504.09
14	637	436	17.21	18.19	-2,046.10	-1,827.61	-21,110.79	7,660.94
15	1,199	969	17.21	18.19	-2,652.30	-3,092.83	-12,978.68	9,871.61
16	1,447	1,285	17.21	18.19	-1,753.90	-2,816.43	-2,576.79	6,172.98
17	1,619	1,494	17.21	18.19	-343.38	-1,780.51	-15.45	2,121.97
18	1,650	1,610	17.21	18.19	1,300.04	-308.76	807.06	59.21
19	764	1,129	17.21	18.19	1,365.96	912.49	4,366.43	737.50
20	649	996	17.21	18.19	1,809.35	1,800.99	14,063.02	3,256.60
21	280	570	17.21	18.19	1,060.61	1,600.69	15,217.92	4,495.10
22	222	468	17.21	18.19	1,062.91	1,782.25	24,366.30	6,787.21
23	145	308	17.21	18.19	839.25	1,480.93	28,114.62	7,120.66
24	107	204	17.21	18.19	726.31	1,184.88	33,465.02	6,882.04
25	56	174	17.21	18.19	436.12	1,184.63	26,451.48	8,065.24
26	38	83	17.21	18.19	333.94	648.08	25,789.31	5,060.38
27	13	47	17.21	18.19	127.24	413.99	12,190.25	3,646.49

28	4	15	17.21	18.19	43.15	147.12	5,021.94	1,443.02
29	1	2	17.21	18.19	11.79	21.62	1,637.98	233.64
SUMAS	9,293	10,017	413.09	436.60	0.00	0.00	83,146.29	82,002.60

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: *Encuesta Nacional de Juventud, 2000.* Instituto Mexicano de la Juventud.

**Se refiere a la media para los hombres, los cálculos se dejan al lector.

**Se refiere a la media para las mujeres, los cálculos se dejan al lector.

Con los datos del Cuadro I.5-34, el cálculo de la desviación estándar

para los hombres es: $s_H = \sqrt{\frac{83,146.29}{9,293}} = \sqrt{8.94719574} = 2.99$, y la desviación

estándar para las mujeres es: $s_M = \sqrt{\frac{82,002.60}{10,017}} = \sqrt{8.186343} = 2.86$.

El coeficiente de variación para los hombres que contestaron a qué edad

habían tenido su primera relación sexual es: $V = \frac{2.99}{17.21} = 0.1737$, mientras que el

coeficiente de variación para las mujeres es: $V = \frac{2.86}{18.19} = 0.1572$. Debido a que

la dispersión en el caso de las mujeres es menor que la dispersión en el caso de los hombres, y por el contrario, la media de los hombres es menor a la media de las mujeres, el coeficiente de variación de las mujeres es menor al coeficiente de variación de los hombres. Por lo anterior, se dice que los hombres entrevistados iniciaron su actividad sexual a edades más diversas que las mujeres a las que se les aplicó el instrumento, quienes inician su actividad sexual entre los 17 y los 19 años de edad, mayoritariamente.

5.5.8 Un ejemplo general.

Para el ejemplo de la sección 5.3.3.1 se calcularán la amplitud, la amplitud intercuartil, la amplitud semiintercuartil, la desviación media, la desviación estándar, la varianza y el coeficiente de variación (véase Cuadro I.5-35). Por razones de espacio, se omitirán todos los cálculos.

CUADRO I.5-35

¿Cuántos libros tiene en su casa? (medidas de dispersión)*

Medida de dispersión	Valor
Amplitud	994 libros.
Amplitud intercuartil	40 libros.
Amplitud semiintercuartil	20 libros.
Desviación media	53.24 libros.
Desviación estándar	95.50 libros.
Varianza	9,119.89
Coefficiente de variación	0.5744

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Prácticas y Consumo Cultural, 2004. UDES, IISUNAM, CONACULTA.

La interpretación de todos esos datos es: entre el encuestado que tuvo más libros y el que tuvo menos, hay 994 libros de diferencia. A su vez, el 50% de las personas entrevistadas que tienen libros, tiene una diferencia de 40 libros, es decir, la mayor parte de las personas tiene una cantidad de libros muy parecida (en este caso entre 10 y 50), lo que difiere en gran medida de los datos extremos. Además, de la mediana a una distancia de a lo más 20 libros se encuentra la mayor parte de los datos (hay que recordar que la mediana para este caso es de 20 libros). De la media a una distancia de a lo más 53.24 libros (la desviación media) se encuentra la mayor parte de los datos de los encuestados (la media para este caso fue de 54.85 libros), y de la media a una distancia de 95.50 libros (la desviación estándar) se encuentra una gran parte de los datos (más del 92% de los datos). Como la desviación media y la desviación estándar son grandes, también se puede concluir que hay una gran dispersión entre los encuestados con más libros en su casa y los encuestados con menos libros. La varianza, que es el cuadrado de la desviación estándar no tiene una interpretación directa. Por último, el coeficiente de variación resultó bastante grande, indicando una vez más, la gran dispersión existente entre los datos y el promedio de los mismos.

En conclusión, todas las medidas de dispersión (excepto la varianza) son de utilidad si se les sabe interpretar adentro de su propia dimensión. Dependiendo de los objetivos del investigador, la o las medidas de dispersión a ocupar pueden ser diferentes.

5.6 Representación gráfica.

Hasta ahora se tiene una gran variedad de medidas de posición, de tendencia central y de dispersión, en las que su construcción requiere de diferentes grados de complejidad. Una manera de facilitar la lectura de los datos, es la tabulación de los datos, tal y como está escrito en la primera sección de este capítulo. Otra manera de facilitar dicha lectura consiste en graficar a los datos. Las gráficas pueden brindarle al investigador una primera idea de la forma en la que se encuentran los datos que pretende analizar, y con ello, puede empezar su trabajo. La estadística brinda una gran variedad de herramientas gráficas para el análisis, y a continuación se presentan algunas de esas herramientas básicas.

5.6.1 El histograma.

Cuando se tienen variables ordinales o de intervalo, se puede construir un gráfico que consista en la representación de una distribución de frecuencias, o de una distribución porcentual, en la que la frecuencia (o porcentaje) de casos se representa mediante un rectángulo en el que el ancho de su base sea igual al tamaño del intervalo, y la altura sea proporcional a la frecuencia (porcentaje) de casos que se incluyen en esa categoría). El gráfico resultante lleva por nombre histograma. Un algoritmo para la construcción de un histograma es el siguiente:

Paso 1. Ordénense los intervalos de menor a mayor (si se trata de datos sin agrupar, transforme los datos a intervalos).

Paso 2. Si es necesario, reconstruya los intervalos para que todos los intervalos sean de igual longitud, y para que todos los datos estén en uno y sólo uno de los intervalos (éste paso no es necesario, pero es muy recomendable seguirlo).

Paso 3. Como todos los intervalos son de igual longitud, la base de cada intervalo será idéntica, y por ello, las alturas variaran en forma proporcional a la cantidad de datos en cada intervalo. Así, grafique las alturas correspondientes a cada intervalo según su frecuencia absoluta (o relativa, en el caso de que sea una representación de una distribución porcentual).

Veamos dos ejemplos.

5.6.1.1. Ejemplo 1

Se construirá el histograma para el ejemplo de la sección 6.3.3.2. El histograma representará a la distribución de frecuencias.

Paso 1. El Cuadro I.5-12 presenta ya a los intervalos ordenados de menor a mayor.

Paso 2. Pese a que el Cuadro I.5-12 presenta a los intervalos ordenados no son de igual longitud, por ello se deben reconstruir los intervalos para que todos sean de igual longitud (véase Cuadro I.5-36).

CUADRO I.5-36
Delegaciones del Distrito Federal de acuerdo a su cantidad de viviendas con techos de materiales precarios*

Intervalo	n_j^*
1. De 0 a 9,999 viviendas**	6
2. De 10,000 a 19,999 viviendas**	5
3. De 20,000 a 29,999 viviendas**	3
4. De 30,000 a 39,999 viviendas**	1
5. De 40,000 a 49,999 viviendas**	0
6. De 50,000 a 59,999 viviendas**	0
7. De 60,000 a 69,999 viviendas**	1
Total	16

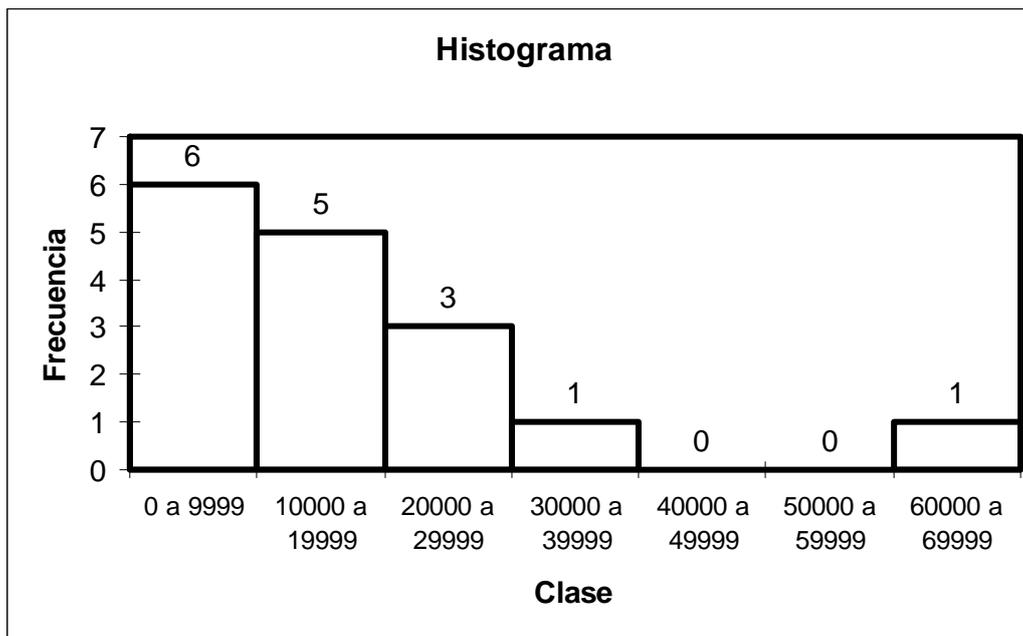
Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000.

**Se refiere a viviendas particulares con techos de materiales ligeros, naturales y precarios.

Paso 3. Con las bases de igual longitud (en este caso, cada base tiene una longitud de 9,999 viviendas), lo que variarán serán las alturas, mismas que deben ser proporcionales a la cantidad de datos en cada intervalo. Por ello, si el intervalo 1 tiene una altura de 6 unidades, el intervalo 2 tendrá una altura de 5 unidades, el intervalo 3 tendrá una altura de 3 unidades, el intervalo 4 tendrá una altura de 1 unidad, el intervalo 5 tendrá una altura de 0 unidades, el intervalo 6 tendrá una altura de 0 unidades, y el intervalo 7 tendrá una altura de 1 unidad (véase Gráfico I.5-1).

GRÁFICO I.5-1



Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000. Los intervalos se refieren a la cantidad de viviendas particulares con techos de materiales ligeros, naturales y precarios se encuentran dentro de ese intervalo. Las frecuencias se refieren a la cantidad de delegaciones en las que el número de viviendas con esa propiedad se encuentra en cada intervalo.

5.6.1.2 Ejemplo 2

Se construirá el histograma para el ejemplo de la sección 5.5.4.3. El histograma representará a la distribución de porcentajes.

Paso 1. El Cuadro I.5-27 tiene ordenados de menor a mayor los datos, sin embargo, estos datos no presentan a los porcentajes, ni son intervalos, por lo que se deben transformar las categorías a intervalos de clase, y se deben obtener los porcentajes (véase Cuadro I.5-37). Para transformar las categorías, se tomará en cuenta que la edad se refiere a años cumplidos, por lo que todas las personas en una categoría dada tienen esa edad con diferente número de días. Para obtener los porcentajes, sólo se toma la frecuencia de cada categoría y se divide entre el total de personas que contestaron a esa pregunta.

CUADRO I.5-37
¿A qué edad tuviste tu primera relación sexual?*

Intervalo	Frecuencia	Porcentaje
De 6 años 0 días a 6 años 364 días	3	0.02%
De 7 años 0 días a 7 años 364 días	3	0.02%
De 8 años 0 días a 8 años 364 días	15	0.08%
De 9 años 0 días a 9 años 364 días	17	0.09%
De 10 años 0 días a 10 años 364 días	18	0.09%
De 11 años 0 días a 11 años 364 días	70	0.36%
De 12 años 0 días a 12 años 364 días	172	0.89%
De 13 años 0 días a 13 años 364 días	391	2.02%
De 14 años 0 días a 14 años 364 días	1,073	5.56%
De 15 años 0 días a 15 años 364 días	2,168	11.23%
De 16 años 0 días a 16 años 364 días	2,732	14.15%
De 17 años 0 días a 17 años 364 días	3,113	16.12%
De 18 años 0 días a 18 años 364 días	3,260	16.88%
De 19 años 0 días a 19 años 364 días	1,893	9.80%
De 20 años 0 días a 20 años 364 días	1,645	8.52%
De 21 años 0 días a 21 años 364 días	850	4.40%
De 22 años 0 días a 22 años 364 días	690	3.57%
De 23 años 0 días a 23 años 364 días	453	2.35%
De 24 años 0 días a 24 años 364 días	311	1.61%
De 25 años 0 días a 25 años 364 días	230	1.19%
De 26 años 0 días a 26 años 364 días	121	0.63%
De 27 años 0 días a 27 años 364 días	60	0.31%
De 28 años 0 días a 28 años 364 días	19	0.10%
De 29 años 0 días a 29 años 364 días	3	0.02%

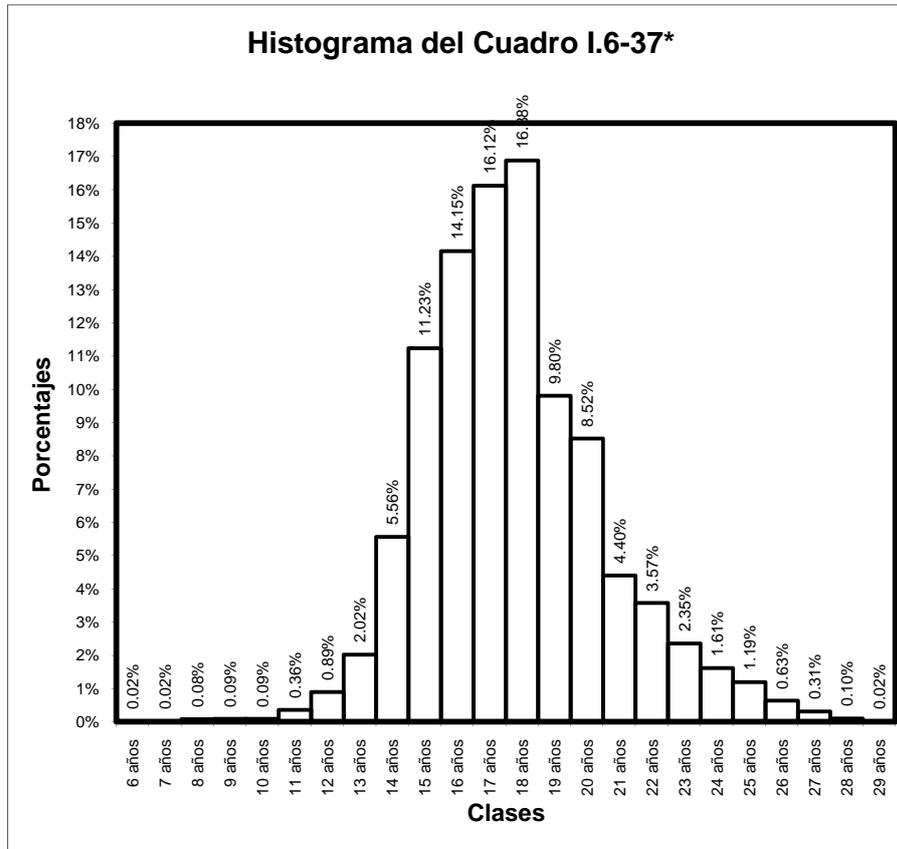
Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Juventud, 2000. Instituto Mexicano de la Juventud.

Paso 2. El Cuadro I.5-37 tiene a los datos ordenados, y cada intervalo tiene la misma longitud (365 días).

Paso 3. Como las bases son de igual longitud, lo único que variará serán las alturas, las cuales deben ser proporcionales a los porcentajes de cada intervalo (véase Gráfico I.5-2). Con todos los datos, se construye el histograma.

GRÁFICO I.5-2



Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Juventud, 2000. Instituto Mexicano de la Juventud. Las clases se refieren a años cumplidos. Los porcentajes representan al porcentaje de personas encuestadas que tuvieron su primera relación sexual en esa edad.

5.6.2 El polígono de frecuencias.

Es un modo muy parecido al histograma para presentar gráficamente a una distribución de frecuencias. Para obtenerlo, se unen los puntos medios de los lados superiores de cada rectángulo por medio de rectas, borrando después a los rectángulos. Un algoritmo es el siguiente:

Paso 1. Constrúyase un intervalo de clase menor al primer intervalo de clase, es decir, conviértase al primer intervalo de clase en el segundo intervalo, construyendo para ello, al intervalo que debería estar antes del primero. Este nuevo primer intervalo tendrá una frecuencia (porcentaje) de cero, ya que no hay ningún dato que esté en ese intervalo.

Paso 2. Constrúyase un intervalo de clase mayor al último intervalo de clase, es decir, conviértase al último intervalo de clase en el penúltimo intervalo, construyendo para ello, al intervalo que debería estar después del último. Este nuevo último intervalo tendrá una frecuencia (porcentaje) de cero, ya que no hay ningún dato que esté en ese intervalo.

Paso 3. Una vez que se tienen todos los intervalos de clase (incluyendo a los de los pasos 1 y 2), constrúyase el histograma.

Paso 4. Tómense los puntos medios de las alturas de los intervalos de clase, y únalos por medio de líneas rectas, uniendo a cada punto únicamente a los dos puntos contiguos (o al único punto contiguo tratándose del primer y del último intervalo).

Paso 5. Bórrense los rectángulos.

5.6.2.1 Ejemplo

A manera de ejemplo ilustrativo, se construirá el polígono de frecuencias, para el histograma de la sección 5.6.1.1

Paso 1. Se debe construir al intervalo que debería estar antes del primer intervalo de clase. En este caso, se trataría del intervalo de -10,000 a -1 viviendas (para que tenga la misma longitud que los intervalos subsecuentes).

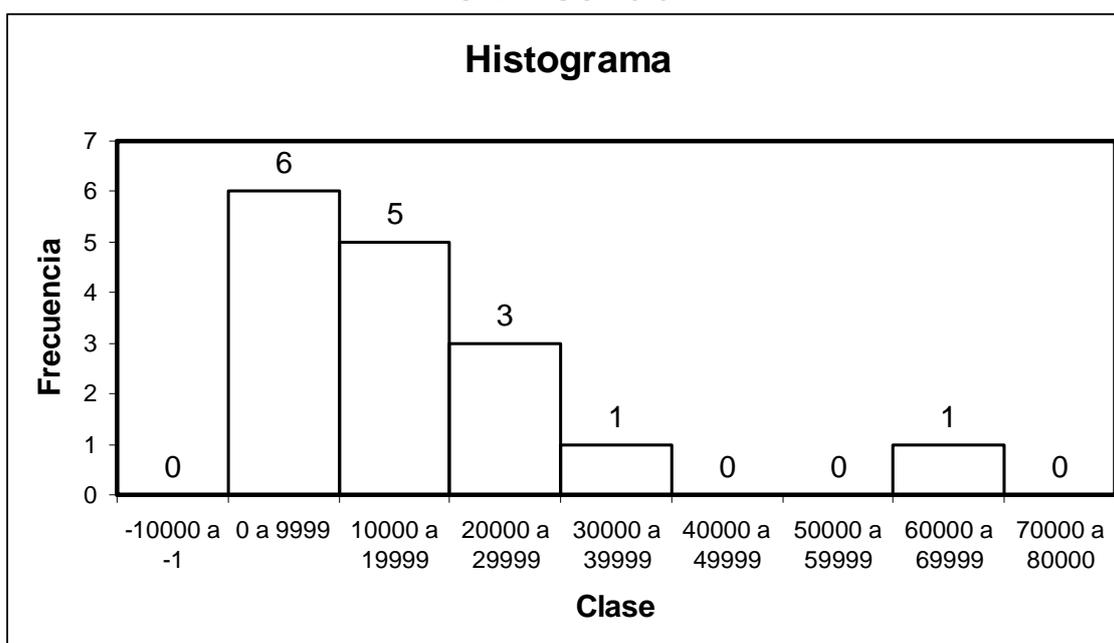
Paso 2. Se debe construir al intervalo que debería estar después del primer intervalo de clase. En este caso, se trataría del intervalo de 70,000 a 79,999 viviendas.

Paso 3 Con esos datos, se construye al nuevo histograma (véase Gráfico I.5-3).

Paso 4 Se calculan los puntos medios de cada intervalo (véase Cuadro I.5-38). Con ello se unen los puntos correspondientes (véase Gráfico I.5-4).

Paso 5. Se borran los rectángulos para que quede únicamente el polígono de frecuencias (véase Gráfico I.5-5).

GRÁFICO I.5-3*



Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000. Los intervalos se refieren a la cantidad de viviendas particulares con techos de materiales ligeros, naturales y precarios se encuentran dentro de ese intervalo. Las frecuencias se refieren a la cantidad de delegaciones en las que el número de viviendas con esa propiedad se encuentra en cada intervalo.

CUADRO I.5-38

Delegaciones del Distrito Federal de acuerdo a su cantidad de viviendas con techos de materiales precarios*

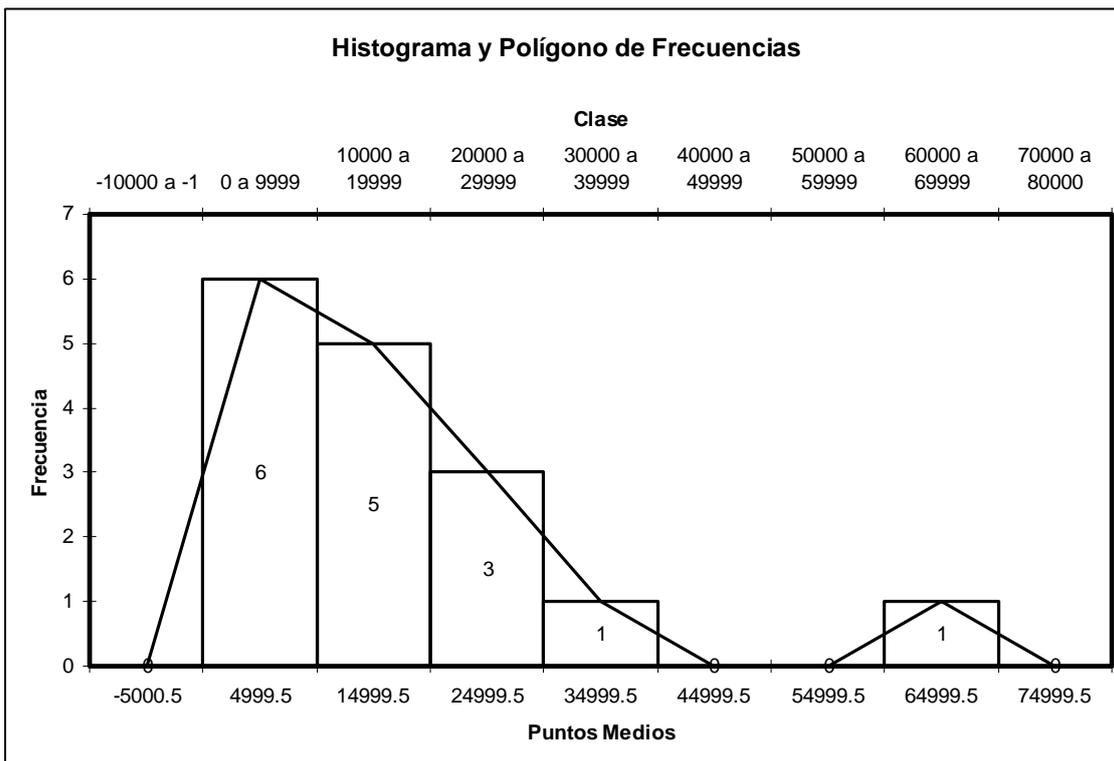
Intervalo	Puntos Medios	n_i *
1. De -10,000 a -1 viviendas**	-5,000.5	0
1. De 0 a 9,999 viviendas**	4,999.5	6
2. De 10,000 a 19,999 viviendas**	14,999.5	5
3. De 20,000 a 29,999 viviendas**	24,999.5	3
4. De 30,000 a 39,999 viviendas**	34,999.5	1
5. De 40,000 a 49,999 viviendas**	44,999.5	0
6. De 50,000 a 59,999 viviendas**	54,999.5	0
7. De 60,000 a 69,999 viviendas**	64,999.5	1
8. De 70,000 a 79,999 viviendas	74,999.5	0
Total		16

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000.

**Se refiere a viviendas particulares con techos de materiales ligeros, naturales y precarios.

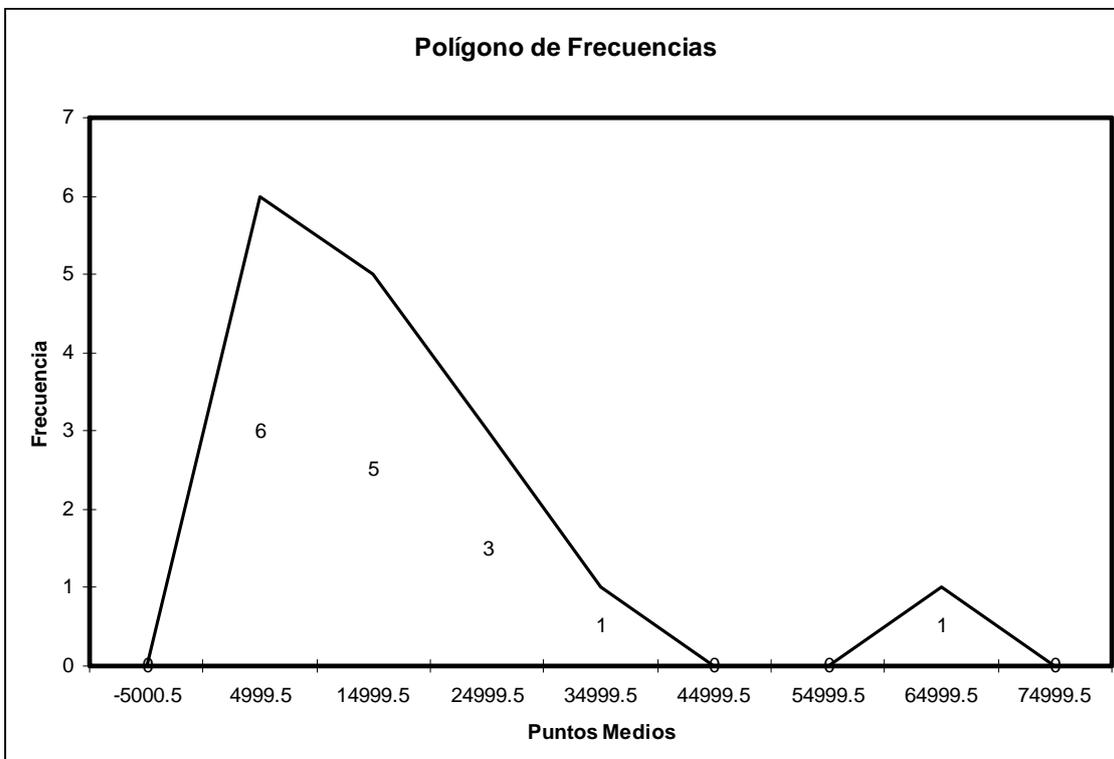
GRÁFICO I.5-4*



Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000. Los intervalos se refieren a la cantidad de viviendas particulares con techos de materiales ligeros, naturales y precarios se encuentran dentro de ese intervalo. Las frecuencias se refieren a la cantidad de delegaciones en las que el número de viviendas con esa propiedad se encuentra en cada intervalo.

GRÁFICO I.5-5*



Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: SCINCE, 2000. Los intervalos se refieren a la cantidad de viviendas particulares con techos de materiales ligeros, naturales y precarios se encuentran dentro de ese intervalo. Las frecuencias se refieren a la cantidad de delegaciones en las que el número de viviendas con esa propiedad se encuentra en cada intervalo.

El algoritmo anterior resulta bastante útil cuando se construyó primero el histograma, o bien, se va a construir. Sin embargo, si no se va a construir el histograma, resulta innecesario construirlo para después construir el polígono de frecuencias. A continuación se da un algoritmo para construir el polígono de frecuencias, sin necesidad de construir el histograma.

Paso 1. Ordénense los intervalos de menor a mayor (si se trata de datos sin agrupar, transforme los datos a intervalos).

Paso 2. Verifique que los intervalos sean de la misma longitud (paso innecesario pero muy conveniente), y que el primero y el último intervalo tengan una frecuencia (porcentaje) de cero. Si los intervalos no cumplen esta propiedad, reconstrúyalos.

Paso 3. Calcule el punto medio de cada intervalo.

Paso 4. Una los puntos medios de los intervalos, por medio de líneas rectas, de tal manera que cada punto quede unido mediante una línea solamente a los dos puntos contiguos (en el caso del primero y del último punto, sólo deben quedar unidos al punto siguiente o al anterior, respectivamente):

5.6.3 La ojiva

La ojiva es una representación gráfica del porcentaje (frecuencia) acumulado hasta cierto punto. Al igual que el polígono de frecuencias, la ojiva se construye a partir de líneas rectas que pasan por los puntos medios de los intervalos. El siguiente es un algoritmo de construcción de la ojiva.

Paso 1. Ordénense los intervalos de menor a mayor (si se trata de datos sin agrupar, transforme los datos a intervalos).

Paso 2. Verifique que los intervalos sean de la misma longitud (paso innecesario pero muy conveniente), y que el primero y el último intervalo tengan un porcentaje (frecuencia) de cero. Si los intervalos no cumplen esta propiedad, reconstrúyalos para que la cumplan.

Paso 3. Calcule el punto medio de cada intervalo.

Paso 4. Calcule el porcentaje (la frecuencia) acumulado hasta un intervalo. Grafique a todos los puntos que tengan como coordenadas a los puntos medios de los intervalos y a los porcentajes acumulados hasta esos intervalos.

Paso 5. Una los puntos del paso 4, por medio de líneas rectas, de tal manera que cada punto quede unido mediante una línea solamente a los dos puntos contiguos (en el caso de los puntos primero y último, sólo deben quedar unidos al punto siguiente o al anterior, respectivamente):

5.6.3.1 Ejemplo

Se construirá la ojiva para el ejemplo de la sección 5.5.4.3. La ojiva representará a la distribución de porcentajes.

Paso 1. El Cuadro I.5-37 muestra a los intervalos ya ordenados de menor a mayor.

Paso 2. Los intervalos son de la misma longitud, pero tanto el primer intervalo como el último tienen un porcentaje mayor a 0. Por lo anterior, se proponen los intervalos “De 5 años 0 días a 5 años 364 días” y “De 30 años 0 días a 30 años 364 días” como intervalos primero y último respectivamente (véase Cuadro I.5-39).

Paso 3. Se deben calcular los puntos medios de cada intervalo, que en este caso corresponderán a cualquier año con 182 días (véase Cuadro I.5-39).

Paso 4. Para calcular el porcentaje acumulado hasta un intervalo se suman los porcentajes de los intervalos anteriores más el del intervalo actual (véase Cuadro I.5-39).

CUADRO I.5-39
¿A qué edad tuviste tu primera relación sexual?*

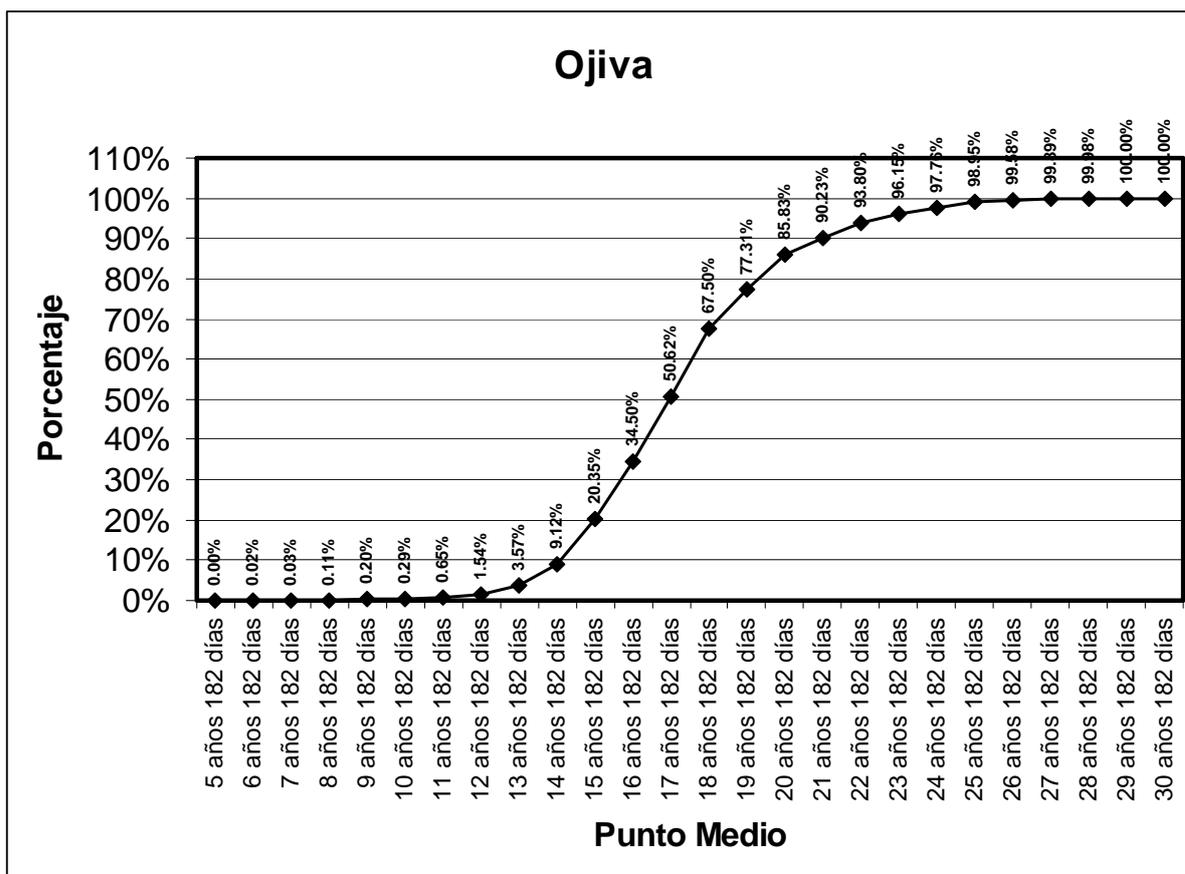
Intervalo	Frecuencia	Porcentaje	Frecuencia Acumulada	Porcentaje Acumulado	Punto Medio
De 5 años 0 días a 5 años 364 días	0	0.00%	0	0.00%	5 años 182 días
De 6 años 0 días a 6 años 364 días	3	0.02%	3	0.02%	6 años 182 días
De 7 años 0 días a 7 años 364 días	3	0.02%	6	0.03%	7 años 182 días
De 8 años 0 días a 8 años 364 días	15	0.08%	21	0.11%	8 años 182 días
De 9 años 0 días a 9 años 364 días	17	0.09%	38	0.20%	9 años 182 días
De 10 años 0 días a 10 años 364 días	18	0.09%	56	0.29%	10 años 182 días
De 11 años 0 días a 11 años 364 días	70	0.36%	126	0.65%	11 años 182 días
De 12 años 0 días a 12 años 364 días	172	0.89%	298	1.54%	12 años 182 días
De 13 años 0 días a 13 años 364 días	391	2.02%	689	3.57%	13 años 182 días
De 14 años 0 días a 14 años 364 días	1,073	5.56%	1762	9.12%	14 años 182 días
De 15 años 0 días a 15 años 364 días	2,168	11.23%	3930	20.35%	15 años 182 días
De 16 años 0 días a 16 años 364 días	2,732	14.15%	6662	34.50%	16 años 182 días
De 17 años 0 días a 17 años 364 días	3,113	16.12%	9775	50.62%	17 años 182 días
De 18 años 0 días a 18 años 364 días	3,260	16.88%	13035	67.50%	18 años 182 días
De 19 años 0 días a 19 años 364 días	1,893	9.80%	14928	77.31%	19 años 182 días
De 20 años 0 días a 20 años 364 días	1,645	8.52%	16573	85.83%	20 años 182 días
De 21 años 0 días a 21 años 364 días	850	4.40%	17423	90.23%	21 años 182 días
De 22 años 0 días a 22 años 364 días	690	3.57%	18113	93.80%	22 años 182 días
De 23 años 0 días a 23 años 364 días	453	2.35%	18566	96.15%	23 años 182 días
De 24 años 0 días a 24 años 364 días	311	1.61%	18877	97.76%	24 años 182 días
De 25 años 0 días a 25 años 364 días	230	1.19%	19107	98.95%	25 años 182 días
De 26 años 0 días a 26 años 364 días	121	0.63%	19228	99.58%	26 años 182 días
De 27 años 0 días a 27 años 364 días	60	0.31%	19288	99.89%	27 años 182 días
De 28 años 0 días a 28 años 364 días	19	0.10%	19307	99.98%	28 años 182 días
De 29 años 0 días a 29 años 364 días	3	0.02%	19310	100.00%	29 años 182 días
De 30 años 0 días a 30 años 364 días	0	0.00%	19310	100.00%	30 años 182 días

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Juventud, 2000. Instituto Mexicano de la Juventud.

Paso 5. El Gráfico I.5-6 ya muestra a la ojiva.

GRÁFICO I.5-6*



Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Juventud, 2000. Instituto Mexicano de la Juventud. Los puntos medios se refieren a los puntos medios de los intervalos en años cumplidos. Los porcentajes representan al porcentaje de personas encuestadas que tuvieron su primera relación sexual en esa edad.

Nótese que el gráfico empieza en el 0%, y termina en el 100%. Tratándose de una ojiva de porcentajes, esta propiedad se cumplirá en todos los casos.

5.7 La sumatoria

Éste último apartado tratará sobre un método abreviado para expresar sumas. Con su uso se puede ahorrar espacio, lo cual redundará en mayor claridad en las expresiones. El símbolo a utilizar es la letra sigma mayúscula (Σ). En la parte de abajo se escribe el índice, así como el límite inferior, y en la parte superior se escribe el límite superior (también se puede escribir el índice nuevamente, pero esto no es usual). Veamos un ejemplo.

Queremos expresar abreviadamente la suma: $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10$. La expresión abreviada, es la siguiente:

$$1+2+3+4+5+6+7+8+9+10 = \sum_{i=1}^{i=10} i = \sum_{i=1}^{10} i.$$

Donde i es el índice, en la parte inferior de la letra sigma se coloca el límite inferior (que indica al primer sumando), y en la parte superior se coloca al límite superior (que indica al último sumando). Nótese que en el primer caso, en la parte superior de la letra sigma también se colocó el índice, mientras que en el segundo ya no se colocó, lo que es lo más común.

El símbolo no sólo es útil para expresar sumas que ya conocemos, sino también es útil para expresar sumas desconocidas, o para generalizar fórmulas. A continuación se presenta con esta notación las fórmulas para la media, y para la desviación estándar.

$$\text{Media} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

$$\text{DesvEst} = s = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Para el cálculo de las desviaciones media y estándar fue necesario tomar valores que pudiéramos garantizar que siempre fueran positivos. Al lector le hubiera podido ser más natural utilizar como medida de desviación a $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})$, dicha fórmula nos ahorraría el cálculo de valores absolutos, o el cálculo de sumas de cuadrados. A continuación se muestra el porqué esa no puede ser una medida de desviación.

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) = \frac{1}{n} \left(\sum_{i=1}^n X_i - \sum_{i=1}^n \bar{X} \right) = \sum_{i=1}^n \frac{1}{n} X_i - \sum_{i=1}^n \frac{1}{n} \bar{X} = \bar{X} - n \frac{1}{n} \bar{X} = \bar{X} - \bar{X} = 0.$$

Ello indica que siempre que se aplica esa fórmula, el resultado es igual a cero. También resulta importante destacar que el uso de éste símbolo, conlleva a una mayor formalidad en las expresiones, y a demostraciones rigurosas.

5.8 BIBLIOGRAFÍA

- [1] **Blalock**, Hubert. Estadística Social. Fondo de Cultura Económica, México, 1966.
- [2] **Croxton**, Frederick; Cowden, Dudley. *Estadística General Aplicada*. Fondo de Cultura Económica, México, 1959.
- [3] **García Ferrando**, Manuel. *Socioestadística, introducción a la estadística en sociología*. Alianza Editoria <<manuales>>, Madrid, España, 1999.
- [4] **Garrido Luque**, Alicia; Álvaro Estramiana, José Luis. *Técnicas de Análisis Estadístico en Ciencias Sociales*. Servicio de publicaciones de la Universidad Complutense, Madrid, España, 1995.
- [5] **Haaser**, Norman; La Salle, Joseph; Sullivan, Joseph. *Análisis Matemático Vol. 1*. Trillas, Segunda Edición, México, 1999.
- [6] **Lobez Urquia**, J; Casa Aruta, E. *Estadística Intermedia*. Editorial vicens – vives, Barcelona, España, tercera edición, 1989.

I.6 Medidas de Concentración

6.1 Introducción

Actualmente se habla mucho sobre la desigualdad existente en la distribución de la riqueza. Así, se llega a decir –por ejemplo- que el 20% de la población mundial obtiene el 80% de los ingresos disponibles. Este tipo de información, regularmente es asumido como verdad, sin ser sometido a algún análisis. Un cuestionamiento que resulta inmediato de la lectura de ese tipo de información es ¿Cómo se llega a ese tipo de conclusiones? La respuesta no es complicada, y para obtenerla, se deben estudiar dos de las medidas de concentración más importantes: la curva de Lorenz y el índice de Gini.

Si bien, las medidas de concentración nos pueden llevar a conclusiones como la expresada en el párrafo anterior, no es lo único que se puede obtener a través de ellas. Como su nombre lo indica, las medidas de concentración nos indican en dónde se concentra la mayor parte de los datos, o, equivalentemente, la distribución de los datos.

En el capítulo anterior, se estudiaron varias formas de ver estas situaciones, sin embargo, para obtener resultados concretos, son necesarios instrumentos más complejos que los ofrecidos desde la estadística descriptiva. Dichos instrumentos radican en poder obtener a partir de la forma de las distribuciones una curva que modele a esos datos. En la parte de la estadística inferencial se verán algunos tipos de curvas que se basan en distribuciones de probabilidad. En este capítulo, sin embargo, se verá cómo se pueden obtener curvas que modelen adecuadamente a la distribución de los datos, sin el uso de la probabilidad, y que son la base de las medidas de concentración que se verán al final del capítulo: La curva de Lorenz y el índice de Gini.

Resulta importante destacar que los métodos aquí mostrados no son los únicos, ni tampoco los mejores, pero tienen la virtud de ser bastante sencillos, y nos arrojan resultados bastante similares a los que pudiéramos obtener con algoritmos de mayor complejidad.

6.2 La interpolación polinomial.

El objetivo de la interpolación consiste en “ajustar” una curva a una serie de puntos. En otras palabras, cuando tenemos datos, y nos gustaría que alguna curva “pasara” por ellos, la interpolación es una herramienta muy utilizada. El término polinomial se agrega al de interpolación cuando queremos que esa curva sea un polinomio.

6.2.1 Los polinomios.

Un polinomio es una función de la forma:

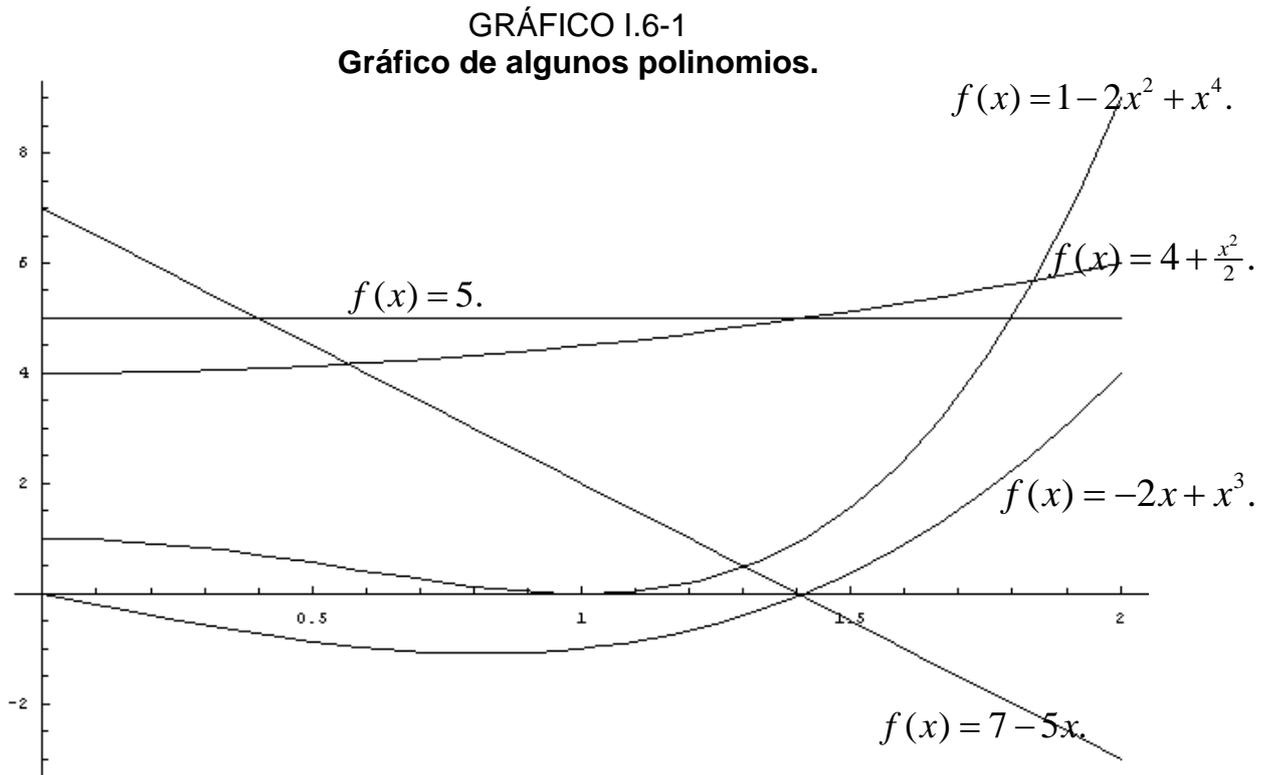
$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \dots + a_nx^n.$$

Donde las a_i 's son valores constantes denominados coeficientes, x representa a la variable, y el exponente indica la potencia a la que es elevada la variable. El polinomio anterior es de grado n , y se dice que el grado de un polinomio es igual al mayor de los exponentes en el que el coeficiente sea diferente de cero. Aunque puede llegar a parecer muy complejo, el manejo de polinomios es bastante simple. De hecho, el lector ya ha trabajado con polinomios desde la enseñanza media. Veamos algunos ejemplos.

1. $f(x) = 5$. Es un ejemplo de una recta constante (con pendiente 0), y es un polinomio de grado 0, ya que el exponente mayor en el que el coeficiente es diferente de cero, es el exponente cero (en efecto, a esta función se le puede ver también como:
 $f(x) = 5 = 5x^0 = 5x^0 + 0x + 0x^2 + 0x^3 + \dots + 0x^n + \dots$).
2. $f(x) = 7 - 5x$. Es un ejemplo de una recta de pendiente distinta de cero. Es un polinomio de grado 1.
3. $f(x) = 4 + \frac{x^2}{2}$. Es un ejemplo de una parábola. Es un polinomio de grado 2, ya que el exponente mayor en el que el coeficiente es diferente de cero, es el exponente 2 (con el coeficiente $1/2$, ya que a esta función se le puede ver también como: $f(x) = 4 + \frac{x^2}{2} = 4x^0 + 0x + \frac{1}{2}x^2$).

4. $f(x) = -2x + x^3$. Es un ejemplo de una curva cúbica. Es un polinomio de grado 3, y nótese que en este caso los coeficientes asociados a las potencias 0 y 2 son iguales a cero.
5. $f(x) = 1 - 2x^2 + x^4$. Es un ejemplo de un polinomio de grado cuatro.

Los ejemplos anteriores, pueden ser vistos en la siguiente gráfica (véase Gráfico I.7-1):



Se pueden demostrar (pero no es objetivo de este texto) los siguientes hechos respecto a los polinomios:

- Por dos puntos distintos pasa una única recta (polinomio de grado 1).
- Por tres puntos distintos y no colineales (es decir, que los tres puntos no se encuentren en una misma recta) pasa una única parábola (polinomio de grado 2).

- Por cuatro puntos distintos y no colineales tres a tres (es decir, que en todos los conjuntos de tres puntos no existe una recta que pase por los tres) pasa un único polinomio de grado 3.
- Por cinco puntos distintos y no colineales tres a tres pasa un único polinomio de grado 4.
- En general, si se tienen $n+1$ puntos distintos y no colineales tres a tres pasa un único polinomio de grado n . Por último, resulta muy importante que dados cualesquiera dos puntos, no estén en una recta vertical, es decir, que dados dos puntos cualesquiera, su primera componente, x , sea distinta.

6.2.2 El algoritmo de interpolación

Cuando tenemos $n+1$ parejas de puntos, de la forma $(x_1, f(x_1)), (x_2, f(x_2)), (x_3, f(x_3)), \dots, (x_{n+1}, f(x_{n+1}))$, y queremos que pase un polinomio de grado n por esos puntos, el algoritmo a seguir es el siguiente:

1. Acomódense las parejas de puntos en un orden vertical. En general, tendremos dos columnas ($j = 2$) con n filas, en la primera aparecerán los x_i , en la segunda aparecerán los $f(x_i)$. Hágase $k = n + 1$.
2. Incrementétese a j en una unidad, i.e., $j = j + 1$.
3. Decrémente a n en una unidad, i.e., $n = n - 1$.
4. La fila i -ésima de la columna j tomará el valor de la fila $i + 1$ -ésima, columna $j - 1$ -ésima, menos el valor de la fila i -ésima, columna $j - 1$ -ésima, dividido entre el valor de la fila $i + j - 2$ -ésima, primera columna, menos el valor de la fila i -ésima, primera columna. Escrito en símbolos:

$$a_{i,j} = \frac{a_{i+1,j-1} - a_{i,j-1}}{a_{i+j-2,1} - a_{i,1}}, \text{ donde } a_{i,j} \text{ representa al valor colocado en la } i\text{-ésima fila, } j\text{-ésima columna. Este paso se sigue para toda } i, \text{ desde que } i \text{ es igual a } 1, \text{ hasta que } i \text{ es igual a } n. \text{ Si alguno de los cocientes es igual a cero, quiere decir que hay dos puntos en una recta vertical, y no es posible obtener el polinomio deseado, termine.}$$

5. Si $n = 1$, pase al paso 6. Si $n > 1$, regrese al paso 2.

6. El polinomio de grado n deseado es:

$$f(x) = a_{1,2} + a_{1,3}(x - a_{1,1}) + a_{1,4}(x - a_{1,1})(x - a_{2,1}) + a_{1,5}(x - a_{1,1})(x - a_{2,1})(x - a_{3,1}) + \dots + a_{1,k+1}(x - a_{1,1})(x - a_{2,1})(x - a_{3,1}) \dots (x - a_{k-1,1})$$

$$= \sum_{h=2}^{k+1} a_{1,h} \prod_{p=1}^{h-2} (x - a_{p,1})$$

Este algoritmo general puede ser muy complicado de seguir. Una manera de facilitar su comprensión y aplicación, es por medio de acomodados triangulares. A continuación se verán dos ejemplos en donde se le da ese acomodo. Se deja al lector verificar que en esos acomodados se siguen todos los pasos del algoritmo.

6.2.2.1 Ejemplo 1

Se desea obtener el polinomio de grado 4 que pasa por los puntos (1,0), (2,3), (0,1), (4,1), y (3,0). En lugar de aplicar directamente al algoritmo, se le dará el acomodo triangular para facilitar su comprensión.

$$\begin{array}{l} \boxed{1|0} \rightarrow \boxed{\frac{3-0}{2-1} = 3} \rightarrow \boxed{\frac{1-3}{0-1} = 2} \rightarrow \boxed{\frac{-\frac{1}{2}-2}{4-1} = -\frac{5}{6}} \rightarrow \boxed{\frac{\frac{5}{6}-(-\frac{5}{6})}{3-1} = \frac{5}{6}} \\ \boxed{2|3} \square \quad \boxed{\frac{1-3}{0-2} = 1} \square \quad \boxed{\frac{0-1}{4-2} = -\frac{1}{2}} \square \quad \boxed{\frac{\frac{1}{3}-(-\frac{1}{2})}{3-2} = \frac{5}{6}} \square \\ \boxed{0|1} \square \quad \boxed{\frac{1-1}{4-0} = 0} \square \quad \boxed{\frac{1-0}{3-0} = \frac{1}{3}} \square \\ \boxed{4|1} \square \quad \boxed{\frac{0-1}{3-4} = 1} \square \\ \boxed{3|0} \square \end{array}$$

Una vez que se tienen esos resultados, el polinomio de grado 4 es:

$$f(x) = 0 + 3(x-1) + 2(x-1)(x-2) - \frac{5}{6}(x-1)(x-2)(x-0) + \frac{5}{6}(x-1)(x-2)(x-0)(x-4).$$

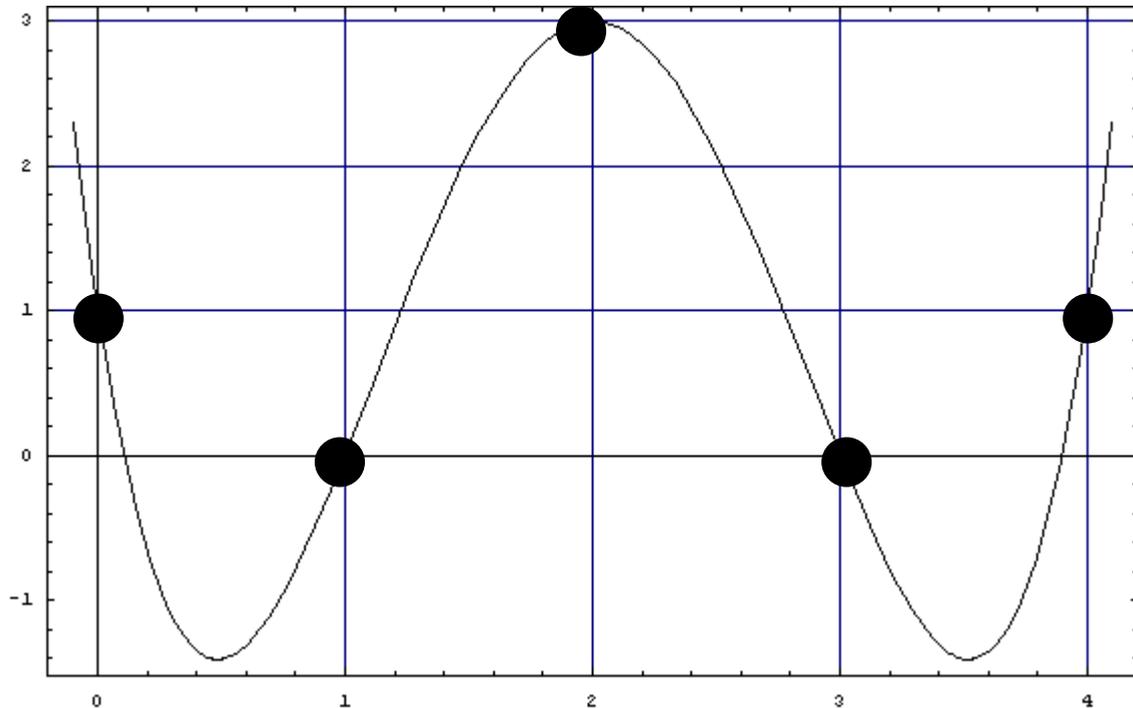
Dicha expresión puede ser simplificada en:

$$1 - \frac{34x}{3} + \frac{97x^2}{6} - \frac{20x^3}{3} + \frac{5x^4}{6}.$$

Y su gráfica es la siguiente (véase Gráfico I.6-2):

GRÁFICO I.6-2

Gráfico del polinomio $1 - \frac{34x}{3} + \frac{97x^2}{6} - \frac{20x^3}{3} + \frac{5x^4}{6}$.



Introducción a la estadística: Un enfoque dirigido a la sociología.

6.2.2.2 Ejemplo 2

Se desea encontrar el polinomio de grado 5 que pasa por los puntos (0,2), (0.2,5), (0.5, 3), (0.8, 1), (0.9,0.3) y (1,0). Se aplicará el algoritmo en su acomodo triangular.

$$\begin{array}{l}
\boxed{0.0} \boxed{2.0} \rightarrow \frac{5-2}{0.2-0} = \boxed{15} \rightarrow \frac{-\frac{20}{3}-15}{0.5-0} = \boxed{-\frac{130}{3}} \rightarrow \frac{0-(-\frac{130}{3})}{0.8-0} = \boxed{\frac{325}{6}} \rightarrow \frac{-\frac{25}{21}-(-\frac{325}{6})}{0.9-0} = \boxed{-\frac{3875}{63}} \rightarrow \frac{\frac{375}{7}-(-\frac{3875}{63})}{1-0} = \boxed{\frac{7250}{63}} \\
\boxed{0.2} \boxed{5.0} \square \quad \frac{3-5}{0.5-0.2} = \boxed{-\frac{20}{3}} \square \quad \frac{-\frac{20}{3}-(-\frac{20}{3})}{0.8-0.2} = \boxed{0} \square \quad \frac{-\frac{5}{6}-0}{0.9-0.2} = \boxed{-\frac{25}{21}} \square \quad \frac{\frac{125}{3}-(-\frac{25}{21})}{1-0.2} = \boxed{\frac{375}{7}} \square \\
\boxed{0.5} \boxed{3.0} \square \quad \frac{1-3}{0.8-0.5} = \boxed{-\frac{20}{3}} \square \quad \frac{-7-(-\frac{20}{3})}{0.9-0.5} = \boxed{-\frac{5}{6}} \square \quad \frac{20-(-\frac{5}{6})}{1-0.5} = \boxed{\frac{125}{3}} \square \\
\boxed{0.8} \boxed{1.0} \square \quad \frac{0.3-1}{0.9-0.8} = \boxed{-7} \square \quad \frac{-3-(-7)}{1-0.8} = \boxed{20} \square \\
\boxed{0.9} \boxed{0.3} \square \quad \frac{0-0.3}{1-0.9} = \boxed{-3} \square \\
\boxed{1.0} \boxed{0.0} \square
\end{array}$$

Una vez que se tienen esos resultados, el polinomio de grado 4 es:

$$\begin{aligned}
f(x) &= 2 + 15(x-0) - \frac{130}{3}(x-0)(x-0.2) + \frac{325}{6}(x-0)(x-0.2)(x-0.5) \\
&\quad - \frac{3875}{63}(x-0)(x-0.2)(x-0.5)(x-0.8) + \frac{7250}{63}(x-0)(x-0.2)(x-0.5)(x-0.8)(x-0.9).
\end{aligned}$$

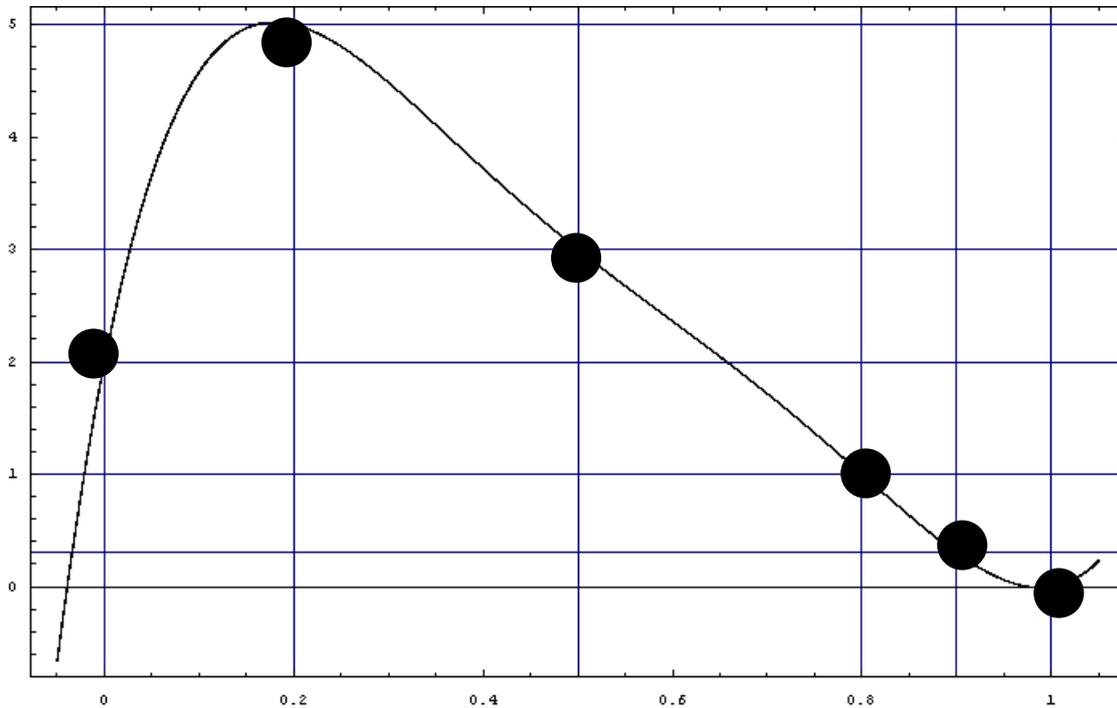
Dicha expresión puede ser simplificada en:

$$f(x) = \frac{1}{252} (504 + 10657x - 50251x^2 + 95190x^3 - 85100x^4 + 29000x^5).$$

Y su gráfica es la siguiente (véase Gráfico I.6-3):

GRÁFICO I.6-3
Gráfico del polinomio

$$f(x) = \frac{1}{252} (504 + 10657x - 50251x^2 + 95190x^3 - 85100x^4 + 29000x^5)$$



Introducción a la estadística: Un enfoque dirigido a la sociología.

6.3 El área de un polinomio entre 0 y 1

Si se tiene un polinomio de grado n , $f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \dots + a_nx^n$, el cual toma sólo valores positivos cuando x es mayor que 0, y menor que 1, su área en esa región es:

$$A = a_0 + \frac{1}{2}a_1 + \frac{1}{3}a_2 + \frac{1}{4}a_3 + \dots + \frac{1}{n+1}a_n = \sum_{i=0}^n \frac{1}{i+1}a_i.$$

La demostración de esa fórmula requiere del principio de inducción, y de operaciones sencillas de cálculo integral, que se omiten para no perder claridad en el presente escrito.

Aplicando la fórmula para el ejemplo de la sección 6.2.2.2 (en la gráfica se puede observar que ese polinomio toma valores positivos entre 0 y 1), el área es:

$$A = \frac{1}{252} \left(504 + \frac{1}{2}10657 - \frac{1}{3}50251 + \frac{1}{4}95190 - \frac{1}{5}85100 + \frac{1}{6}29000 \right) = \frac{693}{252} = 2.75.$$

6.4 La Curva de Lorenz

Es una representación gráfica de inequidad, propuesta por primera vez en 1905 por el estadístico Max Otto Lorenz. En los ejes vertical y horizontal se miden distribuciones de porcentaje acumulado. Es usado en el estudio de la distribución del ingreso y de la concentración industrial, principalmente, aunque puede llegar a tener otros usos.

6.5 El Índice de Gini.

También conocido como Coeficiente de Gini, fue diseñado por el demógrafo Corra de Gini. Es el radio del área entre la curva de Lorenz y la línea de absoluta equidad al área del triángulo bajo la línea.

6.6 Ejemplo 1.

En el cuestionario ampliado del censo del INEGI, aplicado a una muestra de hogares, se preguntó, para cada habitante de un hogar, si percibía alguna fuente de ingreso, y en caso de ser así, se preguntó cuánto dinero obtenía cada persona al mes¹. Utilizando la muestra para el estado de Oaxaca, resultó que del total de personas contempladas en la encuesta (cabe señalar que aunque sólo se aplicaba una encuesta por hogar, se obtenían resultados para todos los individuos del hogar) 113,894 personas percibían algún ingreso. Si se suman los ingresos mensuales de cada persona, se obtiene un total de \$231,528,944, que representaría el ingreso mensual conjunto de las 113,894 personas. Ahora, si se ordena a las personas de acuerdo a su ingreso mensual, de menor a mayor ingreso (véase Cuadro I.6-1), se puede observar la desigualdad existente entre los habitantes de Oaxaca entrevistados, ya que, las primeras 11389 personas (el primer 10%) acumulaba tan solo el 0.79% de los ingresos, mientras que las últimas 11389 personas (el último 10%) acumulaba el 45.22% de los ingresos.

¹ Fuente: XII Censo General de Población y Vivienda, INEGI. Censo ampliado.

CUADRO I.6-1
Distribución de ingresos en Oaxaca en el año 2000*

	Monto	Porcentaje del total	Monto acumulado	Porcentaje acumulado
Persona 1 a persona 11390	\$1,832,229	0.79%	\$1,832,229	0.79%
Persona 11391 a persona 22779	\$4,927,821	2.13%	\$6,760,050	2.92%
Persona 22780 a persona 34168	\$7,440,066	3.21%	\$14,200,116	6.13%
Persona 34169 a persona 45557	\$9,632,219	4.16%	\$23,832,335	10.29%
Persona 45558 a persona 56946	\$12,543,579	5.42%	\$36,375,914	15.71%
Persona 56947 a persona 68335	\$14,655,702	6.33%	\$51,031,616	22.04%
Persona 68336 a persona 79724	\$18,162,587	7.84%	\$69,194,203	29.89%
Persona 79725 a persona 91113	\$23,747,018	10.26%	\$92,941,221	40.14%
Persona 91114 a persona 102502	\$33,883,908	14.63%	\$126,825,129	54.78%
Persona 102503 a persona 113894	\$104,703,715	45.22%	\$231,528,844	100.00%

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: XII Censo General de Población y Vivienda. INEGI. Censo Ampliado. Se refiere únicamente a las personas encuestadas que mencionaron percibir algún ingreso por su trabajo en Oaxaca.

Con estos datos, se puede construir la curva de Lorenz. Para ello, nótese que se sabe que el 0% de la población tiene el 0% del total de los ingresos, es decir, el punto (0,0) está en la curva. Además, el primer 10% acumula el 0.79% del total de los ingresos, es decir, el punto (0.10, 0.0079) está en la curva. Siguiendo con la secuencia, el primer 20% de la población acumula el 2.92% del total de ingresos, es decir, el punto (0.20, 0.0292) está en la curva. El primer 30% de la población acumula el 6.13% del total de ingresos, es decir, el punto (0.30, 0.0613) está en la curva. El primer 40% de la población acumula el 10.29% del total de ingresos, es decir, el punto (0.40, 0.1029) está en la curva. El primer 50% de la población acumula el 15.71% del total de ingresos, es decir, el punto (0.50, 0.1571) está en la curva. El primer 60% de la población acumula el 22.04% del total de ingresos, es decir, el punto (0.60, 0.2204) está en la curva. El primer 70% de la población acumula el 29.89% del total de ingresos, es decir, el punto (0.70, 0.2989) está en la curva. El primer 80% de la población acumula el 40.14% del total de ingresos, es decir, el punto (0.80, 0.4014) está en la curva. El primer 90% de la población acumula el 54.78% del total de ingresos, es decir, el punto (0.90, 0.5478) está en la curva. Por último, el 100% de la población acumula el 100% del total de ingresos, es decir, el punto (1.00, 1.00) está en la curva.

El párrafo anterior nos muestra 11 puntos que sabemos que deben estar en la curva de Lorenz. Como no se tiene una fórmula general para el ajuste de esa curva, pero sí se tienen 11 puntos, se le puede ajustar por interpolación polinomial un polinomio de grado 10. Para un polinomio de un grado tan grande resulta poco ilustrativo realizar las operaciones paso a paso, sin embargo, se muestran los resultados de todas las operaciones en el Cuadro I.6-2. Con base en esa información, se obtiene el siguiente polinomio:

$$\begin{aligned}
 f(x) &= 0.0790x \\
 &+ 0.6700x(x-0.1) \\
 &- 0.4333x(x-0.1)(x-0.2) \\
 &+ 0.5417x(x-0.1)(x-0.2)(x-0.3) \\
 &+ 2.5833x(x-0.1)(x-0.2)(x-0.3)(x-0.4) \\
 &- 19.5833x(x-0.1)(x-0.2)(x-0.3)(x-0.4)(x-0.5) \\
 &+ 81.9444x(x-0.1)(x-0.2)(x-0.3)(x-0.4)(x-0.5)(x-0.6) \\
 &- 227.1825x(x-0.1)(x-0.2)(x-0.3)(x-0.4)(x-0.5)(x-0.6)(x-0.7) \\
 &+ 496.8585x(x-0.1)(x-0.2)(x-0.3)(x-0.4)(x-0.5)(x-0.6)(x-0.7)(x-0.8) \\
 &- 276.1243x(x-0.1)(x-0.2)(x-0.3)(x-0.4)(x-0.5)(x-0.6)(x-0.7)(x-0.8)(x-0.9) \\
 \hline
 f(x) &= 0.5038x - 12.5000x^2 + 138.7655x^3 - 766.8283x^4 + 2479.7081x^5 \\
 &- 4924.0656x^6 + 6040.2774x^7 - 4418.1545x^8 + 1739.4179x^9 - 276.1243x^{10}
 \end{aligned}$$

CUADRO I.6-2
Operaciones para interpolar a una ecuación de grado 10*

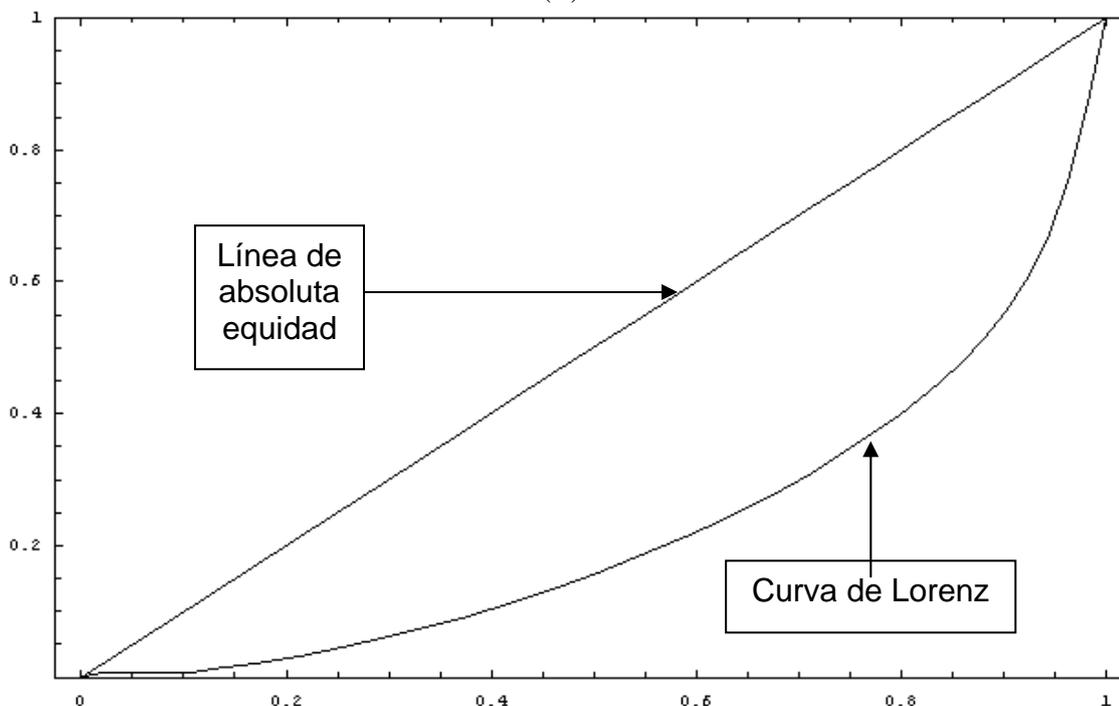
0.0000	0.0000	0.0790	0.6700	-0.4333	0.5417	2.5833	-19.5833	81.9444	-227.1825	496.8585	-276.1243
0.1000	0.0079	0.2130	0.5400	-0.2167	1.8333	-9.1667	37.7778	-99.8016	219.9901	220.7341	
0.2000	0.0292	0.3210	0.4750	0.5167	-2.7500	13.5000	-32.0833	76.1905	418.6508		
0.3000	0.0613	0.4160	0.6300	-0.5833	4.0000	-5.7500	21.2500	411.1111			
0.4000	0.1029	0.5420	0.4550	1.0167	1.1250	7.0000	309.0278				
0.5000	0.1571	0.6330	0.7600	1.4667	4.6250	192.4167					
0.6000	0.2204	0.7850	1.2000	3.3167	100.8333						
0.7000	0.2989	1.0250	2.1950	43.6500							
0.8000	0.4014	1.4640	15.2900								
0.9000	0.5478	4.5220									
1.0000	1.0000										

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Fuente de los datos originales (primeras dos columnas): XII Censo General de Población y Vivienda. INEGI. Censo Ampliado. Se refiere únicamente a las personas encuestadas que mencionaron percibir algún ingreso por su trabajo en Oaxaca.

GRÁFICO I.7-4
Gráfico del polinomio

$f(x) = 0.5038x - 12.5000x^2 + 138.7655x^3 - 766.8283x^4 + 2479.7081x^5 - 4924.0656x^6 + 6040.2774x^7 - 4418.1545x^8 + 1739.4179x^9 - 276.1243x^{10}$, y la recta $f(x) = x$



Introducción a la estadística: Un enfoque dirigido a la sociología.

Una vez que se tiene la fórmula se pueden obtener la curva de Lorenz y el Índice de Gini:

- La curva de Lorenz es el gráfico entre 0 y 1 de la fórmula obtenida de la interpolación (véase Gráfico I.6-4). Aquí es conveniente volver a señalar que este método otorga sólo una curva aproximada a la curva de Lorenz.

- El índice de Gini se obtiene dividiendo el área formada por la intersección del triángulo de absoluta equidad y la curva de Lorenz, entre el área formada por el triángulo de absoluta equidad. En ambos casos, entre 0 y 1. De acuerdo a lo visto en 6.3, el área bajo la curva de Lorenz entre 0 y 1 es:

$$\frac{\frac{1}{2}0.5038 - \frac{1}{3}12.5000 + \frac{1}{4}138.7655 - \frac{1}{5}766.8283 + \frac{1}{6}2479.7081 - \frac{1}{7}4924.0656 + \frac{1}{8}6040.2774 - \frac{1}{9}4418.1545 + \frac{1}{10}1739.4179 - \frac{1}{11}276.1243}{0.225919}$$

- Por otro lado, el área del triángulo de absoluta equidad es: $A = \frac{bh}{2} = \frac{(1)(1)}{2} = \frac{1}{2} = 0.5$.
- Con lo anterior, el índice de Gini es: $I_{Gini} = \frac{0.5 - 0.225919}{0.5} = 0.5482$. Cabe señalar que mientras más se acerque el índice de Gini a 0, mayor será la equidad existente, mientras que si el índice de Gini se acerca a 1, la equidad existente será inferior. En este sentido, se puede comprobar que, en este caso, la inequidad existente entre los trabajadores de Oaxaca consultados es muy grande. Así, mientras que el 20% de la población acumulaba cerca del 60% del total de los ingresos percibidos (59.85%), el 80% de la población restante acumulaba poco más del 40% del total de los ingresos percibidos (40.15%), lo cual es reflejado por un índice de Gini alejado del 0.

En este ejemplo se vio una aplicación del índice de Gini y la curva de Lorenz a la distribución de ingresos. En el siguiente ejemplo se verá su uso aplicado a datos de otra índole.

6.7 Ejemplo 2

En la Encuesta Nacional de Prácticas y Consumo Cultural, levantada por la Unidad de Estudios Sobre la Opinión del Instituto de Investigaciones Sociales de la UNAM, conjuntamente con el Consejo Nacional para la Cultura y las artes durante abril de 2004, se realizó la siguiente pregunta: ¿Cuántos libros tiene en su casa? En el Cuadro I.5-14 se presenta la distribución de los casos. Ahora, si se ordena a las personas de acuerdo al número de libros en su casa, de menor a mayor número de libros (véase Cuadro I.6-3), se puede observar la desigualdad existente, ya que, las primeras 340 personas (el primer 10%) acumulaba tan solo el 0.46% del total de libros, mientras que las últimas 340 personas (el último 10%) acumulaba el 50.68% del total de los libros.

CUADRO I.6-3
Distribución de libros en el hogar mexicano*

Porcentaje personas		Libros	Acumulado de libros	Porcentaje	Porcentaje acumulado
10%	De la persona 1 a la persona 341	854	854	0.46%	0.46%
20%	De la persona 342 a la persona 682	2,016	2,870	1.08%	1.54%
30%	De la persona 683 a la persona 1023	3,340	6,210	1.79%	3.32%
40%	De la persona 1024 a la persona 1364	4,371	10,581	2.34%	5.66%
50%	De la persona 1365 a la persona 1705	6,988	17,569	3.74%	9.40%
60%	De la persona 1706 a la persona 2046	9,558	27,127	5.12%	14.52%
70%	De la persona 2047 a la persona 2387	14,168	41,295	7.58%	22.10%
80%	De la persona 2388 a la persona 2728	18,572	59,867	9.94%	32.05%
90%	De la persona 2729 a la persona 3069	32,270	92,137	17.27%	49.32%
100%	De la persona 3070 a la persona 3406	94,683	186,820	50.68%	100.00%

Introducción a la estadística: Un enfoque dirigido a la sociología.

*FUENTE: Encuesta Nacional de Prácticas y Consumo Cultural. UDES0, IISUNAM, CONACULTA, 2004.

Con los datos anteriores es posible construir la curva de Lorenz, a partir de la interpolación polinomial. Realizar todas las operaciones resultaría poco ilustrativo, por lo que se muestran los pasos en forma abreviada en el Cuadro I.6-4. Con los datos obtenidos, se obtiene la fórmula siguiente para el polinomio de interpolación:

$$\begin{aligned}
 f(x) &= 0.0460x \\
 &+ 0.3100x(x-0.1) \\
 &+ 0.1333x(x-0.1)(x-0.2) \\
 &- 0.9167x(x-0.1)(x-0.2)(x-0.3) \\
 &+ 10.0000x(x-0.1)(x-0.2)(x-0.3)(x-0.4) \\
 &- 42.2222x(x-0.1)(x-0.2)(x-0.3)(x-0.4)(x-0.5) \\
 &+ 135.7143x(x-0.1)(x-0.2)(x-0.3)(x-0.4)(x-0.5)(x-0.6) \\
 &- 368.7996x(x-0.1)(x-0.2)(x-0.3)(x-0.4)(x-0.5)(x-0.6)(x-0.7) \\
 &+ 981.3161x(x-0.1)(x-0.2)(x-0.3)(x-0.4)(x-0.5)(x-0.6)(x-0.7)(x-0.8) \\
 &- 1864.2526x(x-0.1)(x-0.2)(x-0.3)(x-0.4)(x-0.5)(x-0.6)(x-0.7)(x-0.8)(x-0.9)
 \end{aligned}$$

$$\begin{aligned}
 f(x) &= 1.4536x - 38.5928x^2 + 418.6920x^3 - 2405.5871x^4 + 8257.5354x^5 \\
 &- 17761.6922x^6 + 24143.5262x^7 - 20120.5352x^8 + 9370.4528x^9 - 1864.2526x^{10}
 \end{aligned}$$

CUADRO I.7-4
Operaciones para interpolar a una ecuación de grado 10*

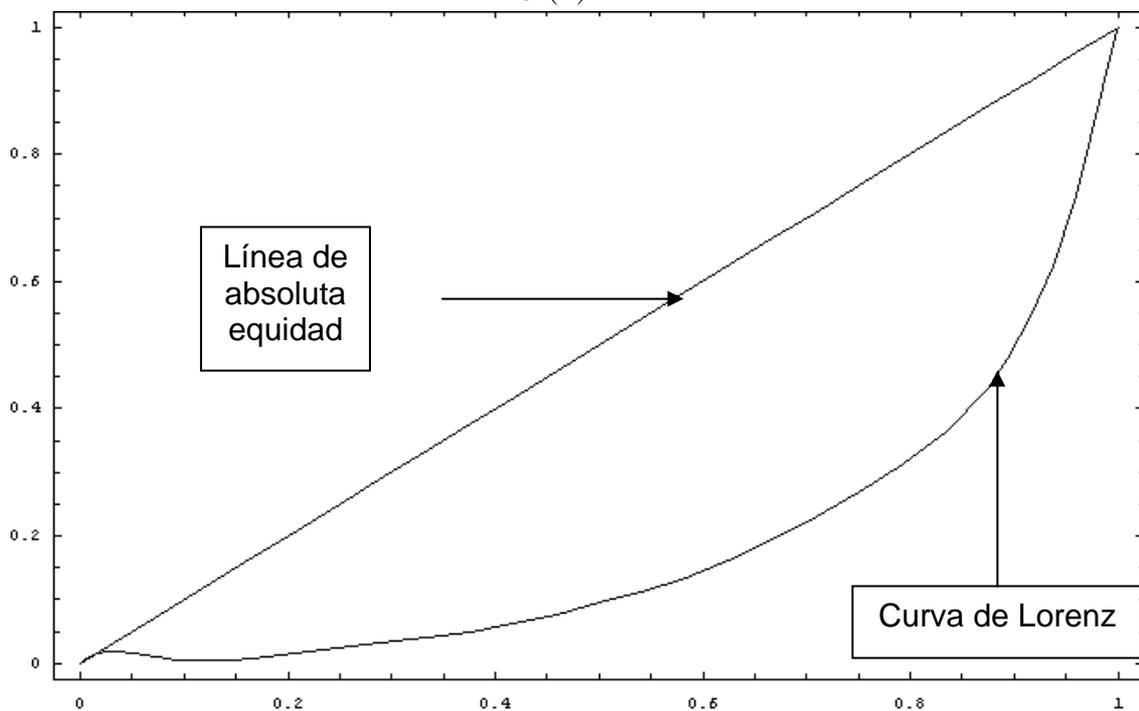
0.0000	0.0000	0.0460	0.3100	0.1333	-0.9167	10.0000	-42.2222	135.7143	-368.7996	981.3161	-1,864.2526
0.1000	0.0046	0.1080	0.3500	-0.2333	4.0833	-15.3333	52.7778	-159.3254	514.3849	-882.9365	
0.2000	0.0154	0.1780	0.2800	1.4000	-3.5833	16.3333	-58.7500	252.1825	-280.2579		
0.3000	0.0332	0.2340	0.7000	-0.0333	4.5833	-18.9167	117.7778	27.9762			
0.4000	0.0566	0.3740	0.6900	1.8000	-4.8750	51.7500	137.3611				
0.5000	0.0940	0.5120	1.2300	-0.1500	21.0000	134.1667					
0.6000	0.1452	0.7580	1.1850	8.2500	88.0833						
0.7000	0.2210	0.9950	3.6600	43.4833							
0.8000	0.3205	1.7270	16.7050								
0.9000	0.4932	5.0680									
1.0000	1.0000										

Introducción a la estadística: Un enfoque dirigido a la sociología.

*Fuente de los datos originales (primeras dos columnas): *Encuesta Nacional de Prácticas y Consumo Cultural*, UDESO, IISUNAM, CONACULTA, 2004

GRÁFICO I.7-5
Gráfico del polinomio

$$f(x) = 1.4536x - 38.5928x^2 + 418.6920x^3 - 2405.5871x^4 + 8257.5354x^5 - 17761.6922x^6 + 24143.5262x^7 - 20120.5352x^8 + 9370.4528x^9 - 1864.2526x^{10}, \text{ y la recta } f(x) = x$$



Una vez que se tiene la fórmula se pueden obtener la curva de Lorenz y el Índice de Gini:

- La curva de Lorenz es el gráfico entre 0 y 1 de la fórmula obtenida de la interpolación (véase Gráfico I.6-5). Nuevamente se recalca que este método otorga sólo una curva aproximada a la curva de Lorenz.
- El índice de Gini se obtiene dividiendo el área formada por la intersección del triángulo de absoluta equidad y la curva de Lorenz, entre el área formada por el triángulo de absoluta equidad. En ambos casos, entre 0 y 1. Conforme a lo visto en la sección 6.3, el área bajo la curva de Lorenz entre 0 y 1 es:

$$\frac{\frac{1}{2}1.4536 - \frac{1}{3}38.5928 + \frac{1}{4}418.6920 - \frac{1}{5}2405.5871 + \frac{1}{6}8257.5354 - \frac{1}{7}17761.6922 + \frac{1}{8}24143.5262 - \frac{1}{9}20120.5352 + \frac{1}{10}9370.4528 - \frac{1}{11}1864.2526}{0.182908}$$

- Por otro lado, el área del triángulo de absoluta equidad

$$\text{es: } A = \frac{bh}{2} = \frac{(1)(1)}{2} = \frac{1}{2} = 0.5.$$

- Con lo anterior, el índice de Gini es: $I_{Gini} = \frac{0.5 - 0.182908}{0.5} = 0.6342$. Cabe

señalar que mientras más se acerque el índice de Gini a 0, mayor será la equidad existente, mientras que si el índice de Gini se acerca a 1, la equidad existente será inferior. En este sentido, se puede comprobar que, en este caso, la inequidad existente es muy grande, de hecho mayor que la existente en el ejemplo anterior. Así, mientras que el 20% de la población acumulaba cerca del 70% del total de los libros (67.95%), el 80% de la población restante acumulaba poco más del 30% del total de los libros (32.05%), lo cual es reflejado por un índice de Gini cercano a 1, y mayor que el índice de Gini del ejemplo anterior.

6.8 Conclusión

En este apartado se vio una forma de que un polinomio pasara por varios puntos, la interpolación polinomial. Este método no es el mejor (aunque lo es para los objetivos de este capítulo), pero sí es una buena aproximación, y permite obtener índices de desigualdad que pueden ser útiles para mostrar la

inequidad existente. Una mejora de la interpolación polinomial sería la interpolación logarítmica, que no se presenta aquí por simplicidad, pero que lograría mejores aproximaciones.

A través de los polinomios es posible aproximar una curva de Lorenz, y obtener el índice de Gini, lo que permite obtener resultados referentes a la inequidad que pueden ser comparables con diferentes tipos de datos. Debido a que la inequidad se presenta en una gran cantidad de casos, es posible realizar comparaciones con datos que no serían comparables de otra forma (en este capítulo se vio una comparación entre libros e ingresos).

6.9 BIBLIOGRAFÍA

- [1] **Cárdenas** Cotiño, Gustavo A. *Diccionario de Ciencias Económico-Administrativas*. CUCEA, Universidad de Guadalajara, 2ª edición. México, 2000.
- [2] **Croxtón**, Frederick; Cowden, Dudley. *Estadística General Aplicada*. Fondo de Cultura Económica, México, 1959.
- [3] **Gordon**, Robert J. *Macroeconomía*. Compañía Editorial Continental SA de CV. México, 1996.
- [4] **Haaser**, Norman; La Salle, Joseph; Sullivan, Joseph. *Análisis Matemático Vol. 1*. Trillas, Segunda Edición, México, 1999.
- [5] **Leithold**, Louis. *El Cálculo*. Oxford University Press 7ª edición. México, 1998.
- [6] **Lobez Urquia**, J; Casa Aruta, E. *Estadística Intermedia*. Editorial vicens – vives, Barcelona, España, tercera edición, 1989.
- [7] **McAllister**, Harry E. *Elementos de estadística en la economía y los negocios*. ECASA, México, 1987.

I.7. Censos Nacionales y Estadísticas Continuas.

Este último capítulo en el apartado de estadística descriptiva hará referencia a las estadísticas de mayor uso en el país. Éstas estadísticas pueden ser analizadas utilizando las herramientas de la estadística descriptiva señaladas en los capítulos anteriores. El presente capítulo tiene como objetivo mostrar un panorama general de lo que se puede esperar de cada una de ellas.

7.1 Censo General de Población y Vivienda¹.

El Censo General de Población y Vivienda es un instrumento mediante el cual se obtienen los datos referentes a toda la población del país. Es llevado a cabo por el Instituto Nacional de Estadística, Geografía e Informática (INEGI), y es realizado cada 10 años, en años múltiplos de 10 (1980, 1990, 2000, etcétera).

Este censo (y en general todos los censos de población y vivienda) constituye la fuente de información estadística más completa en cuanto al desglose geográfico sobre el cual se apoya el conocimiento de la realidad nacional. Permite a los diversos sectores del país elaborar planes y programas de desarrollo, analizar las condiciones de los asentamientos humanos y realizar diversos tipos de investigaciones, entre otros usos.

México tiene una experiencia censal que en tiempos modernos empezó en 1885. Desde entonces, se ha realizado un censo de población cada diez años, en los años que terminan en cero, con excepción de 1920, año en el cual, debido a la situación política predominante en el país, no se realizó, realizándose entonces en 1921. A partir del censo de 1950, además, se incluyeron variables referentes a las viviendas, convirtiendo a los censos de población en censos de población y vivienda.

¹ El desarrollo de 8.1, salvo que se exprese lo contrario, se basa en la información de: "Síntesis Metodológica del XII Censo General de Población y Vivienda 2000" INEGI, julio de 2003.

Con base en la experiencia previa, en consultas y recomendaciones de organismos internacionales, y en consulta con usuarios, investigadores e instituciones públicas y privadas, el XII Censo General de Población y Vivienda contempló variables de estudio que pudieron ser agrupadas en grandes bloques: vivienda; número de residentes y de hogares; y características demográficas, sociales, educativas y económicas de la población. Las variables para cada categoría fueron:

Vivienda

- Tipo y clase de vivienda.
- Materiales de construcción en paredes, techos y recubrimiento del piso.
- Disponibilidad de espacios: total de cuartos, cuartos dormitorio y cocina.
- Disponibilidad y frecuencia del servicio de agua entubada.
- Disponibilidad y exclusividad de servicio sanitario y conexión de agua.
- Disponibilidad de drenaje y electricidad.
- Combustible utilizado para cocinar.
- Tenencia de la vivienda.
- Eliminación de basura.
- Bienes en la vivienda.
- Número de hogares.

Hogares

- Total de residentes.

- Gasto común.

Población

- Sexo, edad y relación de parentesco de los integrantes del hogar.
- Fecundidad y mortalidad.
- Migración.
- Migración internacional.

Características Sociales

- Étnicas.
- Religión.
- Servicios de salud.
- Discapacidad.
- Estado conyugal.

Características Educativas

- Alfabetismo.
- Asistencia escolar.
- Causa de abandono escolar.
- Nivel de instrucción.
- Antecedente escolar.
- Nombre de la carrera.

Características Económicas

- Condición de actividad.
- Ocupación principal.
- Situación en el trabajo.
- Sector de actividad.
- Ingresos por trabajo.
- Horas trabajadas.
- Prestaciones laborales.
- Lugar de trabajo.
- Otros ingresos.

Se recabaron esos datos para todos los hogares del país, con lo que se obtuvo información de todos los habitantes del país. Los censos son fotografías de un instante, sin embargo, debido a que es de esperarse que no ocurran cambios demográficos drásticos, se puede confiar en sus datos para diversos fines.

7.2 Censo Económico².

Los Censos Económicos son la fuente primaria de información económica más completa de México, en la que se basan muchas otras mediciones: encuestas, estudios de planeación, investigaciones económicas, estudios de mercado, etcétera. Su realización constituye una tarea que da cuenta de las principales variables de la economía nacional: personal ocupado, valor de la producción, valor de los insumos utilizados para producir, activos fijos y otros indicadores.

Sectorialmente, los Censos Económicos cubren las actividades referentes a pesca, minería, manufacturas, electricidad, construcción, comercio, transportes, comunicaciones y servicios públicos y privados. Dichas actividades pueden ser analizadas a nivel individual o bien en agregados por rama, subsector y sector de actividad. La información censal es georreferenciada para permitir al usuario tener datos a nivel nacional, por entidad federativa, municipio, localidad e incluso por grupo de manzanas.

El primer levantamiento de los Censos Económicos se realizó en 1930, año en que se visitaron los 48,500 establecimientos más importantes del sector

² El desarrollo de 8.2, salvo que se exprese lo contrario, se basa en la información de: "Panorama General de los Censos Económicos 1999" INEGI, 1999.

industrial. Por su parte, los Censos Económicos 1999 representan el decimoquinto levantamiento censal, con un cubrimiento de poco más de 3.1 millones de unidades económicas dedicadas a toda la gama de actividades económicas.

Los Censos Económicos 1999 captaron información económica básica de las unidades que se dedican a las siguientes actividades:

- Pesca.
- Minería.
- Industrias Manufactureras.
- Electricidad y Agua.
- Industria de la Construcción.
- Comercio.
- Transportes y Comunicación.
- Servicios Financieros, Inmobiliarios y de Alquiler de Bienes Muebles.
- Servicios Técnicos Profesionales, Personales y Sociales.

Su cobertura fue nacional, y se realizó en dos etapas básicas: la enumeración de establecimientos y el levantamiento censal.

La enumeración de establecimientos se realizó para contar con un directorio actualizado de todos los establecimientos existentes en el país, y se realizó a través de un recorrido ordenado y exhaustivo del territorio nacional, manzana por manzana, tocando puerta por puerta, con el objeto de aplicar una cédula, para obtener los datos generales de los establecimientos, su ubicación geográfica, la actividad principal a la que se dedican, su tamaño y tipo de organización.

El levantamiento censal, a su vez, consistió en la aplicación de un cuestionario específico para cada tipo de actividad, con el fin de obtener información detallada de cada unidad económica. A partir de los resultados de

la Enumeración Integral se determinó qué unidades económicas serían censadas por cada Grupo de Trabajo y qué tipo de cuestionario se levantaría a cada una. Esta etapa se realizó entre el 16 de noviembre de 1998 y el 16 de julio de 1999.

7.3 Censo Agrícola y Ganadero³.

El Censo Agrícola-Ganadero es un proyecto de captación de datos que el INEGI lleva a cabo cada diez años desde 1930, con a finalidad de brindar un panorama del comportamiento que presentan las variables relacionadas con las unidades de producción en México, es decir, la forma en que están organizadas, el uso del suelo, la infraestructura con la que cuentan, la cantidad de personas que participan en los procesos de producción, la tenencia de la tierra, la superficie sembrada, la disponibilidad de agua, la cría y explotación de animales y la tecnología empleada, entre otras.

Las variables captadas en el VII Censo Agrícola-Ganadero están agrupadas en 12 grandes temas, y todas ellas buscan satisfacer la demanda que las instituciones usuarias de información agrícola, pecuaria y forestal realizan, así como las recomendaciones internacionales de instituciones como la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO, por sus siglas en inglés).

Los temas considerados en el VII Censo Agrícola-Ganadero fueron:

1. Características generales.
2. Agricultura.
3. Cría y explotación de animales.
4. Explotación forestal.
5. Invernadero y vivero.
6. Crédito y seguro.
7. Organización.
8. Mano de obra.

³ El desarrollo de 7.3, salvo que se exprese lo contrario, se basa en la información de: "*Síntesis Metodológica descenso Agrícola-Ganadero*" INEGI, julio de 2003

9. Vehículos y tractores.
10. Instalaciones y construcciones.
11. Otros usos del suelo.
12. Destino de la producción.

El VII Censo Agrícola-Ganadero se aplicó a todas las unidades estadísticas que se dedicaran a actividades agrícolas, de cría y explotación forestal en todo el territorio nacional, independientemente de que dichas unidades estuvieran localizadas en áreas rurales o urbanas y de que se encontraran o no en explotación durante el periodo del levantamiento de la información. La información recabada correspondió al año agrícola constituido por los ciclos otoño-invierno 1990-1991 y primavera-verano 1991-1991.

7.4 Anuario Estadístico de la UNAM.

Mejor conocidos como Agenda Estadística UNAM, los anuarios estadísticos de la UNAM son publicaciones anuales que reúnen información cuantitativa sobre el quehacer institucional, con el fin de apoyar las actividades de planeación, evaluación y toma de decisiones. La Agenda Estadística UNAM es publicada por la Dirección General de Planeación de la universidad.

En estas publicaciones, se plasman datos cuantitativos referentes a⁴:

- El Personal académico.
 - Tipo de nombramiento.
 - Nombramientos académicos de asignatura.
 - Nombramientos académicos de carrera.
 - Nombramientos académicos en investigación.
 - Personal académico en el SIN.
 - Proyectos de investigación.
- La Población escolar.
 - Población escolar total.
 - Egresados.

⁴ Se refiere a los datos plasmados en “*Agenda Estadística UNAM 2004*” UNAM, México, 2005.

- Exámenes de grado y profesionales.
- Planes de estudio.
- Títulos expedidos.
- El Presupuesto.
 - Presupuesto de ingresos.
 - Presupuesto de egresos.
- Los Programas e intercambio académico.
 - Reconocimientos y estímulos.
 - Formación académica.
 - Desarrollo académico.
 - Educación continua.
 - Acciones de intercambio académico.
 - Servicio social.
- La Infraestructura.
 - Bibliotecas y material bibliográfico.
 - Área construída.
 - Capacidad instalada.
 - Becas para estudiantes.
 - Centros e institutos de investigación.
 - Planteles de educación superior.
 - Planteles de educación media superior.
- La Difusión Cultural.
- El Personal Administrativo.

7.5 Anuario Estadístico del INEGI⁵.

El INEGI realiza dos programas que en forma importante utilizan los resultados obtenidos por los censos nacionales, las encuestas y las estadísticas continuas. El primero es el de Estadística Derivada, por medio del cual se genera el Sistema de Cuentas Nacionales de México. El segundo es el de Integración de Estadísticas, programa orientado a conjuntar en un solo producto estadísticas seleccionadas producidas por el Instituto y por

⁵ El desarrollo de 8.5, salvo que se exprese lo contrario, se basa en la información de: "Anuario Estadístico de los Estados Unidos Mexicanos. Edición 2002" INEGI, México, 2003.

instituciones de los sectores público, privado y social, a fin de facilitar a los usuarios el acceso a un significativo volumen de información sobre un sector o actividad económica, o bien, cubriendo una amplia gama de temas, pero referidos a una unidad geográfica determinada.

A este programa de integración de estadísticas corresponde el Anuario Estadístico de los Estados Unidos Mexicanos, cuyo objetivo es integrar y divulgar información estadística sobre la magnitud, estructura y comportamiento de una gran diversidad de indicadores referidos a los fenómenos económicos y sociales de México, que sirva como base informativa para el conocimiento y análisis de la situación del país.

Sus objetivos específicos son:

- Ofrecer estadística relevante sobre el contexto nacional, bajo criterios uniformes de manejo y presentación.
- Aportar elementos para evaluar la información disponible.
- Contribuir a conformar una base de datos nacional.

La importancia de los anuarios estadísticos de los Estados Unidos Mexicanos del INEGI se sustenta en que propicia el establecimiento de flujos continuos de información estadística entre las instituciones generadoras y los usuarios de la misma; en que presenta bajo un enfoque de complementariedad, estadísticas oficiales y de fuentes privadas; en que contribuye al desarrollo del sistema estadístico nacional y al fortalecimiento del servicio público de información; y en que apoya al fomento de la cultura estadística.

Por último cabe señalar que el INEGI realiza varios anuarios estadísticos, referidos a diferentes ámbitos. Uno de ellos es el Anuario Estadístico de las Entidades Federativas, el cual trae la información de forma independiente para cada una de las 32 entidades federativas.

7.6 Anuario Estadístico de la ONU⁶.

El Anuario Estadístico de la ONU (Statistical Yearbook en inglés, y Annuaire Statistique en francés), tiene como principal objetivo agrupar en un solo volumen una compilación de las estadísticas internacionales disponibles sobre condiciones y actividades sociales y económicas, a nivel mundial, regional y por nación, cubriendo, cuando es posible, un periodo de diez años.

La mayor parte de las estadísticas presentadas en el anuario son extraídas de bases de datos más detalladas y especializadas que son preparadas por la División de Estadística (Statistics Division), y por muchas otras agencias internacionales de estadística. Así, mientras las publicaciones especializadas se concentran en el monitoreo de algunos temas concretos sobre un campo social o económico específico, el Anuario Estadístico provee de datos en un ámbito más general, describiendo las estructuras sociales y económicas, sus cambios y actividades. Su objetivo ha consistido en coleccionar, sistematizar y coordinar los componentes más esenciales de información estadística comparable, que pueda otorgar un panorama general de los procesos sociales y económicos.

El contenido del Anuario Estadístico se realiza para que sea útil a cualquier lector. El Anuario provee información para diversos órganos de las Naciones Unidas, para organismos internacionales, para organizaciones gubernamentales y no gubernamentales, para órganos estadísticos, económicos y sociales de las naciones, para instituciones científicas y educacionales, para bibliotecas y para el público en general. Sus datos también son de interés para compañías y empresas dedicadas a la investigación de mercado. El Anuario Estadístico provee información sistemática en un rango bastante amplio de tópicos sociales y económicos que son de interés para la propia organización, para sus miembros, y para los habitantes del orbe. A su vez, facilita el análisis de los diferentes tópicos al sistematizar y coordinar datos

⁶ El desarrollo de 8.6, salvo que se exprese lo contrario, se basa en la información de: "Statistical Yearbook Forty-eighth issue" United Nations, USA, 2004.

de diferentes campos del conocimiento. Entre los diferentes temas que abarca, destacan:

- Crecimiento global económico y las condiciones económicas relacionadas.
- La situación económica de los países en desarrollo y el progreso frente a objetivos adoptados por las Naciones Unidas para el desarrollo de los mismos entre una década y otra.
- Población y urbanización, su crecimiento e impacto.
- Producción de energía, su consumo, y el desarrollo de nuevas fuentes de energía.
- Expansión del comercio.
- Alimentación y el alivio de la hambruna.
- Educación, entrenamiento y erradicación del analfabetismo.
- Mejora de las condiciones generales de vida.
- Contaminación y protección del medio ambiente.
- Asistencia a los países en desarrollo para propósitos sociales y económicos.

7.7 Registro Nacional de Población.

En México, corresponde a la Dirección General del Registro de Población e Identificación Personal, dependiente de la Secretaría de Gobernación, registrar y acreditar fehacientemente la identidad de las personas que integran la población del país y a los mexicanos que radican en el extranjero, a efecto de otorgarles certeza jurídica para el ejercicio pleno de sus derechos.

Por lo anterior, es el Registro Nacional de Población el encargado de tener los registros correspondientes a las estadísticas continuas. Por medio de él es posible saber la cantidad de nacimientos, emancipaciones, matrimonios, divorcios y fallecimientos ocurridos en un periodo de tiempo específico.

7.8 BIBLIOGRAFÍA.

- *“Agenda Estadística UNAM 2004”* UNAM, México, 2005.
- *“Anuario Estadístico de los Estados Unidos Mexicanos. Edición 2002”*. INEGI, México, 2003.
- *“Panorama General de los Censos Económicos 1999”* INEGI, México, 2003.
- *“Síntesis Metodológica del Censo Agrícola-Ganadero”* INEGI, México, 2003.
- *“Síntesis Metodológica del XII Censo General de Población y Vivienda 2000”*. INEGI, México, 2003.
- *“Statistical Yearbook. Forty-eighth issue”* United Nations, USA, 2004.

I.8 Conclusiones generales

- Las ciencias sociales paulatinamente han aumentado los análisis que realizan con base en datos recopilados por diversas fuentes. Por ello, los futuros sociólogos requieren una base de conocimientos estadísticos sólida, es decir, deben conocer la amplia gama de medidas que utiliza la estadística descriptiva, y que les puede ayudar a analizar y sintetizar sus resultados.
- Resulta de gran importancia desarrollar un libro de texto que coadyuve con la enseñanza de la estadística a los estudiantes de sociología, pues si bien, en una primera instancia, el aprendizaje está a cargo de los profesores en las aulas, el alumno que conozca un libro de texto en el que confíe y que requiera del uso de la estadística, tendrá mayores posibilidades de éxito frente al alumno que, perdido en la amplia bibliografía, aplique técnicas que no sean las más adecuadas a lo que está haciendo.
- El texto presenta gran cantidad de ejemplos que ilustran las diversas técnicas existentes. La mayor parte de los ejemplos está basada en estadísticas recopiladas recientemente. Con ello se gana claridad en la aplicación de las herramientas estadísticas, a la vez que se mantiene al alumno en contacto con resultados de interés nacional. Sin embargo, una desventaja de éste método es que el texto se desactualizará en poco tiempo. Así, la escritura de un libro con éstas características requiere que el autor esté comprometido con su actualización permanente.
- Aunque el tema no es desarrollado por esta tesis, es necesario que el alumno se familiarice con algún software estadístico. Probablemente dicho software pueda calcular todas las medidas estadísticas presentadas aquí. El alumno que habiendo leído un texto como este, utilice el software, tendrá la ventaja de saber lo que quiere y puede esperar del software, evitando errores.

- Aun cuando no se descarta la pertinencia de varios de los temas aquí tratados (y algunos no incluidos en esta tesis, pero sí contemplados dentro del plan de estudios), varios de ellos no están directamente relacionados con el estudio de la estadística, y podrían ser incluidos en un curso de matemáticas en general, pues al tratarlos como parte de un curso de estadística, el alumno se puede confundir, y no se logra la especificidad requerida para el aprendizaje óptimo de la materia.
- Los cuadros de doble entrada serán de utilidad para el alumno que tenga datos cuyo análisis se facilite al presentarlos simultáneamente, como podría ser el caso del sexo y la edad. El aprendizaje de los principales métodos para acomodar dichos cuadros le permitirá construir cuadros que puedan ser entendidos por la mayoría de las personas, a la vez que él podrá interpretar cuadros de doble entrada con facilidad al saber la forma en que comúnmente son puestos los datos.
- El término de conjunto, como un ente matemático, permite clasificar adecuadamente al universo en estudio de acuerdo a características de sus elementos. El estudiante que comprenda los conceptos básicos será capaz de lograr abstracciones matemáticas que, además, serán de utilidad al momento de estudiar algunos conceptos de probabilidad que son útiles para el aprendizaje y uso de la estadística inferencial, la cual es teorizada a partir de distribuciones de probabilidad.
- Los números índice resumen en un solo número gran cantidad de información. El alumno que se familiarice con su construcción podrá analizarlos con precisión, e incluso, construir indicadores que le puedan ser de utilidad en sus investigaciones. Para ello, es importante que entienda que varios de los índices que existen en la actualidad se construyen con la finalidad de comparar la realidad en periodos de tiempo anteriores con el periodo actual. Para ejemplificar estos puntos se presentan los índices de precio y cantidad, lo que también introduce al alumno en algunos conceptos básicos de la teoría económica.

- La gran cantidad de ejemplos ilustra todos los conceptos concernientes a la estadística descriptiva, lo que permitirá que, al término de la lectura del capítulo 5, el alumno conozca los conceptos que le permitirán en un futuro describir bases de datos.
- Ahondando en la teoría económica, las medidas de concentración permitirán al estudiante analizar conceptos como pobreza y desigualdad. Como parte de los objetivos de esta tesis, se presentan casos que ilustran el hecho de que las medidas de desigualdad no tienen que ser ocupadas exclusivamente en el campo económico, sino que además, tienen aplicaciones en otros campos donde se analiza la desigualdad presente entre segmentos de la sociedad.
- Al final del trabajo se presenta una breve descripción sobre algunos censos y encuestas que levanta INEGI. Con ello se espera que el estudiante sepa que existe una gran cantidad de datos a los que puede recurrir cuando realice ciertas investigaciones.
- Finalmente, esta tesis le presenta al alumno de sociología una alternativa de estudio, y le muestra que los conceptos estadísticos le serán útiles a lo largo de su desarrollo profesional.

BIBLIOGRAFIA

1. “*Agenda Estadística UNAM 2004*” UNAM, México, 2005.
2. “*Anuario Estadístico de los Estados Unidos Mexicanos. Edición 2002*”. INEGI, México, 2003.
3. “*Panorama General de los Censos Económicos 1999*” INEGI, México, 2003.
4. “*Síntesis Metodológica del Censo Agrícola-Ganadero*” INEGI, México, 2003.
5. “*Síntesis Metodológica del XII Censo General de Población y Vivienda 2000*”. INEGI, México, 2003.
6. “*Statistical Yearbook. Forty-eighth issue*” United Nations, USA, 2004.
7. **Babbie**, Earl. *The practice of social research*. Belmont, CA. USA, 10ª edición, 2004.
8. **Blalock**, Hubert M. Jr. *Estadística Social*. Fondo de Cultura Económica. México, 1966.
9. **Cárdenas** Cotiño, Gustavo A. *Diccionario de Ciencias Económico-Administrativas*. CUCEA, Universidad de Guadalajara, 2ª edición. México, 2000.
10. **Cárdenas**, Humberto, et al. *Álgebra Superior*. Trillas, 2ª edición. México, 1999.
11. **Cramer**, Duncan. *Basic statistics for social research*. Routledge London and New York. UK, 1997.
12. **Croxtan**, Frederick E. *Estadística General Aplicada*. Fondo de Cultura Económica, México, 1959.
13. **Enderton**, Herbert. *Elements of Set Theory*. Academic Press inc, New York, 1977.
14. **Fox**, Karl A. *Manual de Econometría*. Amorrorto editores. Argentina, 1973.
15. **García Ferrando**, Manuel. *Socioestadística Introducción a la Estadística en Sociología*. Alianza Editorial Manuales / Ciencias Sociales. España, 1999.

16. **Garrido Luque**, Alicia, et. al. *Técnicas de Análisis Estadístico en Ciencias Sociales*. Servicio de Publicaciones Universidad Complutense. España, 1995.
17. **Gordon**, Robert J. *Macroeconomía*. Compañía Editorial Continental SA de CV. México, 1996.
18. **Haaser**, Norman B, et al. *Análisis Matemático*, volumen 1. Trillas, 2ª edición. México, 1999.
19. **Leithold**, Louis. *El Cálculo*. Oxford University Press 7ª edición. México, 1998.
20. **Lobez Urquia**, J; Casa Aruta, E. *Estadística Intermedia*. Editorial vicens – vives, Barcelona, España, tercera edición, 1989.
21. **Mangin, Jean** – Pierre Lévy; **Varela Malloy**, Jesús; **Abad González**, Julio. *Análisis Multivariable para las Ciencias Sociales*. Pearson Educación, España, 2004.
22. **McAllister**, Harry E. *Elementos de estadística en la economía y los negocios*. ECASA, México, 1987.
23. **Nidditch**, P. H. *El desarrollo de la Lógica matemática*. Ediciones Cátedra, Col. Teorema, 5ª edición. España, 1995.
24. **Salas**, Javier. *Econometría Aplicada a los países en desarrollo. El caso mexicano*. FCE, México, 1990.
25. **Scharles**, Maurice, et. al. *Economic Análisis, Theory and Applications*. Irwin inc, 6a edición. Estados Unidos, 1992.
26. **Werner**, Lloyd. *Social Class in America*. Science research associates, inc, Chicago, USA, 1949.